2013-10-03

The Serum Resistome of a Globally Disseminated Multidrug Resistant Uropathogenic <italic>Escherichia coli</italic> Clone

Phan, MD

http://hdl.handle.net/10026.1/4514

10.1371/journal.pgen.1003834
PLoS Genet
Public Library of Science

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.
The Serum Resistome of a Globally Disseminated Multidrug Resistant Uropathogenic *Escherichia coli* Clone

Minh-Duy Phan\(^1,2\), Kate M. Peters\(^1,2\), Sohinee Sarkar\(^1,2\), Samuel W. Lukowski\(^3\), Luke P. Allsopp\(^1,2\), Danilo Gomes Moriel\(^1,2\), Maud E. S. Achard\(^1,2\), Makrina Totsika\(^1,2\), Vikki M. Marshall\(^3\), Mathew Upton\(^4\), Scott A. Beatson\(^1,2\), Mark A. Schembri\(^1,2\*)

\(^1\) Australian Infectious Diseases Research Centre, The University of Queensland, Brisbane, Queensland, Australia, \(^2\) School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland, Australia, \(^3\) Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia, \(^4\) Faculty of Medicine and Dentistry, University of Plymouth, Plymouth, United Kingdom

\* Current address: Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland.

**Abstract**

*Escherichia coli* ST131 is a globally disseminated, multidrug resistant clone responsible for a high proportion of urinary tract and bloodstream infections. The rapid emergence and successful spread of *E. coli* ST131 is strongly associated with antibiotic resistance; however, this phenotype alone is unlikely to explain its dominance amongst multidrug resistant uropathogens circulating worldwide in hospitals and the community. Thus, a greater understanding of the molecular mechanisms that underpin the fitness of *E. coli* ST131 is required. In this study, we employed hyper-saturated transposon mutagenesis in combination with multiplexed transposon directed insertion-site sequencing to define the essential genes required for *in vitro* growth and the serum resistome (i.e. genes required for resistance to human serum) of *E. coli* EC958, a representative of the predominant *E. coli* ST131 clonal lineage. We identified 315 essential genes in *E. coli* EC958, 231 (73%) of which were also essential in *E. coli* K-12. The serum resistome comprised 56 genes, the majority of which encode membrane proteins or factors involved in lipopolysaccharide (LPS) biosynthesis. Targeted mutagenesis confirmed a role in serum resistance for 46 (82%) of these genes. The murein lipoprotein Lpp, along with two lipid A-core biosynthesis enzymes WaaP and WaaG, were most strongly associated with serum resistance. While LPS was the main resistance mechanism defined for *E. coli* EC958 in serum, the enterobacterial common antigen and colanic acid also impacted on this phenotype. Our analysis also identified a novel function for two genes, *hyxA* and *hyXR*, as minor regulators of O-antigen chain length. This study offers novel insight into the genetic make-up of *E. coli* ST131, and provides a framework for future research on *E. coli* and other Gram-negative pathogens to define their essential gene repertoire and to dissect the molecular mechanisms that enable them to survive in the bloodstream and cause disease.


**Editor:** Paul M. Richardson, MicroTrek Incorporated, United States of America

**Received** April 11, 2013; **Accepted** August 12, 2013; **Published** October 3, 2013

**Copyright:** © 2013 Phan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by a grant from the Australian National Health and Medical Research Council [APP1012076]. MAS was supported by an Australian Research Council (ARC) Future Fellowship [FT100100662]. SAB was supported by an ARC Australian Research Fellowship [DP0801347]. MT was supported by an ARC Discovery Early Career Researcher Award [DE130101169]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: m.schembri@uq.edu.au

**Introduction**

*Escherichia coli* O25b:H4-ST131 (*E. coli* ST131) is a recently emerged, globally disseminated clone that is often multidrug resistant and is responsible for a high proportion of community-acquired urinary tract and bloodstream infections [1–6]. *E. coli* ST131 strains are also capable of causing complicated infections including acute pyelonephritis, osteomyelitis, septic arthritis and septic shock [7,8]. *E. coli* ST131 are commonly associated with production of the CTX-M-15 enzyme, currently the most widespread extended spectrum β-lactamase (ESBL) of its type in the world [1,9]. In addition to resistance against oxyimino-cephalosporins (i.e. cefotaxime, ceftazidime), and monobactams, *E. coli* ST131 strains are often co-resistant to fluoroquinolones [3,10]. Indeed, most fluoroquinolone-resistant *E. coli* strains belong to a recently emerged and dominant subgroup of ST131 strains [11]. Some *E. coli* ST131 strains have also been reported to produce carbapenemases [12–14], thus severely limiting treatment options that are currently available against this clinically predominant clone [15].

*E. coli* ST131 strains, like many other uropathogenic *E. coli* (UPEC) strains, are derived from phylogenetic group B2 [3]. Typically, UPEC strains possess a large and diverse range of virulence factors that contribute to their ability to cause urinary tract and bloodstream infections, including adhesins, toxins, siderophores and protectins [15,16]. Several studies have demonstrated that *E. coli* ST131 strains possess a similar suite of virulence factors and cause invasive disease, leading to the hypothesis that...
Author Summary

The emergence and rapid dissemination of new bacterial pathogens presents multiple challenges to healthcare systems, including the need for rapid detection, precise diagnostics, effective transmission control and effective treatment. *E. coli* ST131 is an example of a recently emerged multidrug resistant pathogen that is capable of causing urinary tract and bloodstream infections with limited available treatment options. In order to increase our molecular understanding of *E. coli* ST131, we developed a high-throughput transposon mutagenesis system in combination with next generation sequencing to test every gene for its essential role in growth and for its contribution to serum resistance. We identified 315 essential genes, 270 of which were conserved among all currently available complete *E. coli* genomes. Fifty-six genes that define the serum resistome of *E. coli* ST131 were identified, including genes encoding membrane proteins, proteins involved in LPS biosynthesis, regulators and several novel proteins with previously unknown function. This study therefore provides an inventory of essential and serum resistance genes that could form a framework for the future development of targeted therapeutics to prevent disease caused by multidrug-resistant *E. coli* ST131 strains.

The widespread pathogenic success of *E. coli* ST131 strains may be in part due to enhanced virulence [7,8,17]. However, it has become clear from recent studies that *E. coli* ST131 strains do not possess a heightened virulence potential compared to other UPEC or B2 *E. coli* strains in causing invasive infections [18] or infections in nematodes and zebrafish embryos [19]. Thus, other factors such as enhanced metabolic capacity have been proposed to contribute to the fitness and pathogenic success of this dominant clone [20,21].

The genome sequence of one of the best-characterized *E. coli* ST131 strains, EC958, was recently described [22]. *E. coli* EC958 is a member of the pulsed-field gel electrophoresis (PFGE) defined UK epidemic strain A, which represents one of the major pathogenic lineages (PFGE strains A–E) of ESBL-producing *E. coli* causing urinary tract infections (UTI) across the UK [23]. *E. coli* EC958 is resistant to eight antibiotic classes, including oxyiminocephalosporins, fluoroquinolones and sulphonamides. *E. coli* EC958 colonizes the bladder of mice in a type 1 fimbiae-dependent manner [22], can invade into bladder epithelial cells and form intracellular bacterial communities, can establish both acute and chronic UTI [24] and can inhibit the contraction of ureters, *in vitro* [25].

The ability to resist the bactericidal activity of serum, and thus survive in the bloodstream, represents an essential virulence trait for UPEC and other extra-intestinal *E. coli* (ExPEC) strains, including *E. coli* ST131 [26–28]. In *E. coli*, several mechanisms have been shown to contribute to serum resistance. The importance of O-antigens and K capsules in resistance to serum has been recognized since the 1960s and 1980s, respectively [29–31]; and their multiple types, combinations and length contribute differently to serum resistance [32–35]. The major outer membrane protein OmpA [36], plasmid-encoded proteins TraT [37,38] and Iss [39], and the phage membrane protein Bor [40] have also been reported to contribute to serum resistance in *E. coli*. Notably, each of these resistance mechanisms has been studied in isolation and in different strain backgrounds. Thus, while serum resistance is clearly a complex phenotype determined by multiple elements, little is known about the combination of factors that contribute to resistance in a single strain.

High-throughput transposon mutagenesis combined with genome-wide targeted sequencing was used recently to study the essential genes in *Salmonella enterica* serovar Typhi and *Caulobacter crescentus* [41,42]. Langridge et al. also used their transposon directed insertion-site sequencing (TraDIS) method to assay every gene for its role in the survival of *S. Typhi* in the presence of bile salts [42]. Similar approaches (INSeq, HITS, Tri-seq) have also been applied to a range of organisms to study gene requirements for survival in particular niches [43–46]. Here, we adapted TraDIS and designed a multiplexing method to define the essential genes required for *in vitro* growth (i.e. Luria-Bertani agar media supplemented with 30 μg/ml Cm at 37°C) and the serum resistome (i.e. genes required for resistance to human serum, in *E. coli* EC958). We show that the essential gene list of *E. coli* EC958 comprises 315 genes, 231 of which are shared with *E. coli* K-12. We also define for the first time a comprehensive inventory of genes required for resistance to human serum. Our study provides a molecular blueprint for understanding the mechanisms employed by *E. coli* ST131 to survive, grow in the bloodstream and cause disease.

Results

Application of multiplexed TraDIS to the *E. coli* ST131 strain EC958

Approximately 1 million mutants were generated in the *E. coli* ST131 strain EC958 [22] using an in-house miniTn5 transposon carrying a chloramphenicol (Cm) resistance gene derived from the pKD3 plasmid [47]. A primer comprising four functional regions was designed to facilitate specific screening of the transposon insertion sites on the Illumina HiSeq 2000 platform while allowing for intra-lane multiple sample indexing (Figure S1). This primer contained (5’-3’): (i) the P5 sequence to bind to TruSeq flowcells, (ii) the Illumina read 1 sequencing primer binding site, (iii) a 6-bp index sequence for multiple sample barcoding and (iv) a 25-bp transposon specific sequence designed to amplify the last 12 bp of the transposon and its adjacent genomic sequence. Using this custom primer, we successfully sequenced the transposon insertion sites for 6 samples on both TruSeq version 2 and version 3 flowcells (Figure 1A). Each sample yielded from 6.8 million to 15 million reads that were tagged with transposon specific sequence, 71% of which were reliably mapped to EC958 draft chromosome (excluding unscaffolded contigs and plasmids) (Table 1). All experiments were performed in duplicate, with the correlation coefficient for the number of insertions per gene for each pair of samples close to 1 (R²=0.99) and thus demonstrating a high level of reproducibility for each experiment (Figure 1B).

Essential genes in *E. coli* EC958

We initially used our saturated random insertional mutant library to determine the ‘essential genes of EC958’, defined as the set of genes required for growth on LB agar supplemented with Cm 30 μg/ml. We extracted genomic DNA (in duplicates: input A and B) directly from the library pool and sequenced using our multiplexed TraDIS protocol. We combined the reads from input A and B to maximize the coverage resulting in 16 million transposon-tagged reads, of which 11 million uniquely mapped to the EC958 chromosome, resulting in 502,068 unique insertion sites. This equates to an average of one insertion site every 9.92 bp, with a very low probability of having 100 consecutive bp without interruption by chance (P=4.2×10⁻⁸).
The essential gene list was identified using a statistical analysis similar to that described by Langridge et al., which recognized two distinct distributions of insertion indexes (number of insertions divided by gene length) for non-essential genes (gamma) and essential genes (exponential) and called those with insertion indexes less than or equal to the intercept of the two distributions as essential [42]. In our data, an insertion index cut-off of 0.0158, resulted in the identification of 315 genes as essential (Table S1). This cut-off is equivalent to a log2-likelihood-ratio (LLR) of $2^{3.6}$, which means that our essential genes are at least 12 times more likely to belong to the exponential distribution (essential) than the gamma (non-essential) distribution.

The functional category of each gene was identified based on the COG (Clusters of Orthologous Groups) numbers from the EC958 annotation (accession number PRJEA61443). Figure 2 shows an overview of essential functions in EC958 compared with the total number of genes in each functional category. Genes involved in translation, ribosomal structure and biogenesis account for 25% of the total number of essential genes in EC958, which is 42% of the total number of genes in this category. The second most abundant category in the essential gene list comprised genes involved in cell wall/membrane/envelope biogenesis (12%), followed by genes involved in coenzyme transport and metabolism. There were 23 essential genes with functions not identified in the COG database.

To investigate the conservation of EC958 essential proteins among different E. coli pathotypes, we performed blastx alignment (FASTA v36) between the essential protein sequences and translated DNA from 50 E. coli complete genomes (Table S2). There were 270 (86%) proteins conserved across all genomes investigated. An additional 17 proteins were also present in more than 90% of the genomes. Only 6 proteins were specific for EC958 (not found in 50 genomes) (Table S3).

The serum resistome of E. coli EC958

Saturated transposon mutagenesis in combination with next-generation sequencing is a powerful tool for whole genome, high-throughput identification of all candidate genes involved in a particular phenotype. Here, we used our transposon mutant library in combination with TraDIS to identify genes from EC958 involved in resistance to human serum, thus enabling us to define the serum resistome of EC958. We designed a mutant selection procedure in which 1 million mutants were exposed to pooled fresh human serum for 90 minutes and then allowed to grow in LB broth for 4 hours before genomic DNA extraction. This procedure permitted the growth of serum resistant mutants while inhibiting mutants that were sensitive to serum (Figure 3A). The procedure was performed in parallel with control samples where fresh serum was replaced by inactivated serum that lacked bactericidal activity (data not shown). The genomic DNA from test and control samples were sequenced using our modified Illumina multiplexed TraDIS procedure (Figure 3B) to generate multiple datasets (Figure 1A) that were analysed by the Bioconductor package edgeR after filtering out genes identified as essential [48].

The serum resistance genes were identified as genes that have significant reduction in read counts in the test samples compared to the control samples (i.e. more mutants survived after serum treatment) (Table S4). A stringent threshold of log2 fold change of read counts (logFC) less than $-1$ and an adjusted p-value less than 0.05 was used.

<table>
<thead>
<tr>
<th>Tagged reads</th>
<th>Reads mapped to EC958 chromosome (%)</th>
<th>Unique insertion sites in EC958</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input A &amp; B</td>
<td>16,281,107</td>
<td>11,558,154 (70.99)</td>
</tr>
<tr>
<td>Test A</td>
<td>6,819,691</td>
<td>4,879,098 (71.54)</td>
</tr>
<tr>
<td>Test B</td>
<td>7,572,730</td>
<td>5,408,136 (71.41)</td>
</tr>
<tr>
<td>Ctrl A</td>
<td>8,350,687</td>
<td>5,918,766 (70.87)</td>
</tr>
<tr>
<td>Ctrl B</td>
<td>15,006,139</td>
<td>10,727,954 (71.49)</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pgen.1003834.t001
0.001 was used to identify significant genes that are involved in serum resistance (Figure S2). Figure 4 shows the names and genomic locations of the 56 genes that satisfied these stringent criteria.

Twenty-two (39.3%) of the genes belong to three operons responsible for LPS biosynthesis (including both O-antigen biosynthesis and lipid A-core biosynthesis) and enterobacterial common antigen (ECA) biosynthesis. This result represents the first layer of evidence demonstrating the importance of the O25 antigen as well as ECA in an E. coli ST131 background for protection from the bactericidal activity of human serum. Detailed characterisation of the O25 antigen gene cluster is discussed in subsequent sections. ECA is common to all Enterobacteriaceae and is expressed in both serum resistant (smooth) and sensitive (rough) strains, except for rough strains that are defective in the shared biosynthesis pathway affecting both O-antigen and ECA [49,50]. Seven out of 12 genes in the ECA operon were required for serum resistance as determined by the TraDIS technique.

The remaining 60.7% of genes in the serum resistome identified by TraDIS included genes encoding lipoprotein, membrane proteins, regulators and hypothetical proteins (Table 2). Some of these also affect LPS such as sfaH (EC958_4322), encoding a known regulator required for LPS biosynthesis [51,52] and virulence of pathogenic E. coli strains [53,54], whilst others represent genes that have not previously been shown to be associated with serum resistance. The murein lipoprotein gene lpp (EC958_1897) showed the greatest difference between the test and control samples (logFC of −10), followed by pgm (EC958_0806), encoding phosphoglucomutase. Four hypothetical proteins were also identified, two of which (EC958_0460 and EC958_0461) were further characterized in this study (see below).

Validation of serum resistance genes
As mentioned above, we employed a stringent threshold combining fold change and statistical significance to define the set of 56 genes in the EC958 serum resistome. In order to validate these findings, we attempted to test all 56 genes independently for their role in serum resistance. Using a modified lambda red mediated-homologous recombination approach [22,47] we successfully generated defined knock-out mutants for 54 genes (96.4%) in EC958. We were unable to obtain mutants for the remaining 2 genes (acrA and EC958_2373) despite multiple attempts (Table 2).
The 54 defined mutants were subjected to serum susceptibility testing, whereby the number of surviving colonies after a 90-minute exposure to fresh pooled human serum was compared to the number of colonies prior to treatment (Table 2). A mutant was defined as susceptible to serum when its log difference was at least 1 (i.e., 10-fold reduction after exposure to serum). Forty-one genes (75.9%) contributed to serum resistance in EC958 using this assay. In the case of the remaining 13 mutants, it is possible that the lack of susceptibility to human serum was a reflection of the assay, suggesting that survival in serum within a mixed population of one million mutants may be very different from survival of a pure population carrying the same defective mutation. Therefore, to better mimic the condition of TraDIS library serum selection, a competitive assay was devised where EC958 wild-type was mixed equally with a mutant before exposure to serum and the competitive index of the mutant was measured. Using this competitive assay, five of the twelve mutants were significantly attenuated compared to the wild-type EC958 strain (Table 2). Thus, the overall number of validated susceptible mutants was 46 out of 54 tested (85.2%).

Additional characteristics of serum resistant mutants

We hypothesized that one mechanism associated with enhanced sensitivity to human serum could be due to decreased membrane integrity caused by destabilization of the outer leaflet of the outer membrane. In order to test this, we examined the survival of the 54 mutants in response to outer membrane stresses (i.e., SDS) and osmotic potential (i.e., NaCl). In total, 50.0% (27/54) of the mutants displayed enhanced sensitivity to SDS and 11.1% (6/54) of the mutants displayed enhanced sensitivity to NaCl (Table 2). A comparative analysis of these phenotypes in the context of serum sensitivity is presented below.

O25b antigen biosynthesis genes conferring resistance to serum

It is well established that O-antigen represents the main determinant for serum resistance in E. coli. However, there are more than 180 different O-antigens that have been defined in E. coli and these may contribute differently to serum resistance in individual bacterial strains [32,33]. Furthermore, direct genetic evidence linking O-antigen biosynthesis to serum resistance is only available for a small number of specific O-antigen types. With this in mind, a detailed characterization of the O25b biosynthesis genes was performed using sequence comparison for function prediction in combination with analysis of LPS composition to deduce the role of each gene in resistance to serum.

Similar to most E. coli strains, the O-antigen biosynthesis operon is located between the gilF and gnd genes in EC958. Figure 5A shows a comparison of the EC958 O-antigen operon with the equivalent operon from the K-12 strain MG1655 and the O25 serotype E. coli strain E47a [32,55].

Nucleotide sugar biosynthesis genes.

The four genes encoding dTDP-α-L-rhamnose biosynthesis enzymes were highly conserved in the three strains, consistent with the presence of α-L-rhamnose in the O16 and O25 antigen repeat unit expressed by MG1655 and EC958, respectively [56–58]. The dTDP-α-L-rhamnose biosynthesis pathway starts from α-D-glucose-1-phosphate to produce dTDP-α-L-rhamnose via the catalysis of RmlA (to make dTDP-α-D-glucose), RmlB (to make dTDP-4-dehydro-6-deoxy-α-D-glucose), RmlC (to make dTDP-4-dehydro-6-deoxy-β-mannose) and RmlD (to make dTDP-α-L-rhamnose), respectively. LPS gel analysis of EC958 mutants lacking one of these four enzymes showed that RmlC and RmlD were required for the biosynthesis of O-antigen and mutation in either gene resulted in cells with only the lipid A core (Figure 5C). On the contrary, the
LPS patterns of \( rmlA \) and \( rmlB \) mutants showed only a small change in intensity but retained most of the O-antigen chain length distribution, except for the very long chain length band. This might be explained by the existence of \( RffH \) and \( RffG \), two isozymes of \( RmlA \) and \( RmlB \), in EC958. \( RffH \) and \( RffG \), however, were not able to fully compensate for \( RmlA \) or \( RmlB \) in resistance to serum (Table 2). The inability to produce the very long chain length O-antigen in \( rmlA \) and \( rmlB \) mutants might be a crucial factor in determining their susceptibility to human serum. Only the \( rmlC \) mutant displayed altered sensitivity to SDS, while mutations in \( rmlABCD \) did not affect sensitivity to NaCl (Table 2).

The O-antigen operon in E47a has another set of nucleotide sugar biosynthesis genes (\( fnlABC \)) for the biosynthesis of UDP-N-acetyl-\( \alpha \)-L-fucosamine (UDP-FucNAc) from UDP-N-acetyl-\( \alpha \)-D-glucosamine (UDP-GlcNAc). Curiously, EC958 does not have these three genes in the operon or anywhere else in its genome.

**O-antigen processing genes.** The O-antigen flippase (\( Wzx \), EC958\_2377) and O-antigen polymerase (\( Wzy \), EC958\_2374) from EC958 are highly similar to the corresponding genes from E47a. In contrast, both genes share very low similarity to the corresponding genes from MG1655. Mutation of the \( wzx \) gene in MG1655 results in the accumulation of high levels of the UndPP-O unit in the cytoplasm [59] and hypersensitivity to several antibiotics and other agents including nalidixic acid, tetracycline, mitomycin C and hydrogen peroxide [60]. Transcriptional analyses of MG1655 responses to the broad-spectrum biocide polyhexamethylene biguanide suggested that \( Wzx \) (also known as RfbX) might be involved in cellular stress responses [61]. In contrast to the previous studies, the \( wzx \) gene was defined as essential in this study (Figure 4, Table S1). This might be explained by the fact that Cm was used to select for transposon mutants and the \( wzx \) mutants might be hypersensitive to this antibiotic. The \( wzx \) O-antigen polymerase gene (EC958\_2374) was also defined as essential in EC958 and thus was not further characterized. Mutation of the chain length regulator gene (\( wzz \), EC958\_2368) showed a non-modal distribution of O-antigen chain length in its LPS pattern (Figure 5C), an observation consistent with previous study describing the role of this gene [62]. The reduction of long chain length O-antigen in this mutant is likely to account for the serum sensitivity shown in our assay (Table 2).

**Glycosyltransferases genes.** Glycosyltransferases (GTs) are required to form the glycosidic bonds between sugars in an O-antigen repeat unit. There are now more than 100,000 GTs within 94 families (http://www.cazy.org). Based on the structure of the O25 repeat unit [63,64], we predicted a requirement for 4 GTs in EC958. Bioinformatic analysis (blastp) confirmed that there are indeed 4 GTs within the O-antigen operon: EC958\_2371, EC958\_2372, EC958\_2375 and EC958\_2376. Surprisingly, two of the predicted GTs in EC958 are very different from those in E47a. Whether this is related to the different source of \( \alpha \)-L-FucNAc in EC958 due to the lack of \( fnlABC \) remains to be
Table 2. Genes required for serum resistance identified by multiplexed TraDIS.

<table>
<thead>
<tr>
<th>EC958 locus_tag</th>
<th>Gene Name</th>
<th>Matched in MG1655</th>
<th>Product</th>
<th>Serum*</th>
<th>LPS*</th>
<th>SDS MIC† (%)</th>
<th>NaCl MIC‡ (M)</th>
<th>COG code</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC958_0147</td>
<td>dnaJ</td>
<td>b0015</td>
<td>DnaJ chaperone</td>
<td>1</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>WT O</td>
</tr>
<tr>
<td>EC958_0152</td>
<td>nhaA</td>
<td>b0019</td>
<td>Sodium/proton antiporter</td>
<td>0b</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>P</td>
</tr>
<tr>
<td>EC958_0260</td>
<td>acnB</td>
<td>b0118</td>
<td>aconitate hydratase</td>
<td>1</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>C</td>
</tr>
<tr>
<td>EC958_0460</td>
<td></td>
<td>0460</td>
<td>-</td>
<td>3</td>
<td>X</td>
<td>WT</td>
<td>WT</td>
<td>-</td>
</tr>
<tr>
<td>EC958_0461</td>
<td></td>
<td>0461</td>
<td>-</td>
<td>1b</td>
<td>X</td>
<td>WT</td>
<td>WT</td>
<td>TK</td>
</tr>
<tr>
<td>EC958_0602</td>
<td>acrA</td>
<td>b0463</td>
<td>multidrug efflux system</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>M</td>
</tr>
<tr>
<td>EC958_0784</td>
<td>nagC</td>
<td>b0676</td>
<td>Transcriptional regulator, repressor of N-acetylglucosamine</td>
<td>1</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>KG</td>
</tr>
<tr>
<td>EC958_0785</td>
<td>nagA</td>
<td>b0677</td>
<td>N-acetylglucosamine-6-phosphate deacetylase</td>
<td>3</td>
<td>X</td>
<td>0.063</td>
<td>WT</td>
<td>G</td>
</tr>
<tr>
<td>EC958_0806</td>
<td>pgm</td>
<td>b0688</td>
<td>Phosphoglucomutase</td>
<td>4</td>
<td>X</td>
<td>0.016</td>
<td>WT</td>
<td>G</td>
</tr>
<tr>
<td>EC958_0856</td>
<td>tolQ</td>
<td>b0737</td>
<td>membrane spanning protein in Tol-Pal cell envelope complex</td>
<td>3</td>
<td>WT</td>
<td>0.031</td>
<td>WT</td>
<td>U</td>
</tr>
<tr>
<td>EC958_0858</td>
<td>tolA</td>
<td>b0739</td>
<td>membrane anchored protein in Tol-Pal cell envelope complex</td>
<td>3</td>
<td>WT</td>
<td>0.031</td>
<td>WT</td>
<td>M</td>
</tr>
<tr>
<td>EC958_0859</td>
<td>tolB</td>
<td>b0740</td>
<td>periplasmic protein in Tol-Pal cell envelope complex</td>
<td>0</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>U</td>
</tr>
<tr>
<td>EC958_0871</td>
<td>galE</td>
<td>b0759</td>
<td>UDP-galactose-4-epimerase</td>
<td>1</td>
<td>X</td>
<td>WT</td>
<td>WT</td>
<td>M</td>
</tr>
<tr>
<td>EC958_1112</td>
<td>1112</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>X</td>
<td>WT</td>
<td>WT</td>
<td>-</td>
</tr>
<tr>
<td>EC958_1114</td>
<td>1114</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>-</td>
</tr>
<tr>
<td>EC958_1897</td>
<td>lpp</td>
<td>b1677</td>
<td>murein lipoprotein</td>
<td>6</td>
<td>WT</td>
<td>0.063</td>
<td>WT</td>
<td>M</td>
</tr>
<tr>
<td>EC958_2368</td>
<td>wzz</td>
<td>b2027</td>
<td>regulator of O-antigen chain length</td>
<td>1</td>
<td>X</td>
<td>WT</td>
<td>WT</td>
<td>M</td>
</tr>
<tr>
<td>EC958_2371</td>
<td>2371</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>LA</td>
<td>0.063</td>
<td>WT</td>
<td>R</td>
</tr>
<tr>
<td>EC958_2373</td>
<td>2373</td>
<td>-</td>
<td>-</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>E</td>
</tr>
<tr>
<td>EC958_2378</td>
<td>rmiC</td>
<td>b2038</td>
<td>dTDP-4-deoxyhydroxynose-3,5-epimerase</td>
<td>4</td>
<td>LA</td>
<td>0.063</td>
<td>WT</td>
<td>M</td>
</tr>
<tr>
<td>EC958_2379</td>
<td>rmiA</td>
<td>b2039</td>
<td>glucose-1-phosphate thymidylyltransferase</td>
<td>2</td>
<td>X</td>
<td>WT</td>
<td>WT</td>
<td>M</td>
</tr>
<tr>
<td>EC958_2380</td>
<td>rmiD</td>
<td>b2040</td>
<td>dTDP-4-dehydroxynose reductase</td>
<td>4</td>
<td>LA</td>
<td>WT</td>
<td>WT</td>
<td>M</td>
</tr>
<tr>
<td>EC958_2381</td>
<td>rmiB</td>
<td>b2041</td>
<td>dTDP-glucose-4,6-dehydratase</td>
<td>3</td>
<td>X</td>
<td>WT</td>
<td>WT</td>
<td>M</td>
</tr>
<tr>
<td>EC958_2390</td>
<td>wcaL</td>
<td>b2050</td>
<td>glycosyl transferase involved in colanic acid synthesis</td>
<td>0b</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>M</td>
</tr>
<tr>
<td>EC958_2391</td>
<td>gmm</td>
<td>b2051</td>
<td>GDP-mannose mannosyl hydrolase</td>
<td>1b</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>F</td>
</tr>
<tr>
<td>EC958_2394</td>
<td>wcaF</td>
<td>b2054</td>
<td>predicted acyl transferase</td>
<td>4</td>
<td>X</td>
<td>0.063</td>
<td>0.5</td>
<td>R</td>
</tr>
<tr>
<td>EC958_2394</td>
<td>amD</td>
<td>b2256</td>
<td>undecaprenyl phosphate-alpha-L-ara4FN deformylase</td>
<td>2</td>
<td>WT</td>
<td>0.063</td>
<td>WT</td>
<td>G</td>
</tr>
<tr>
<td>EC958_2396</td>
<td>amE</td>
<td>b4544</td>
<td>AmE/AmF undecaprenyl-phosphate-alpha-L-Ara4FN flippase</td>
<td>1</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>P</td>
</tr>
<tr>
<td>EC958_2597</td>
<td>amF</td>
<td>b2258</td>
<td>AmE/AmF undecaprenyl-phosphate-alpha-L-Ara4FN flippase</td>
<td>2</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>-</td>
</tr>
<tr>
<td>EC958_2652</td>
<td>dedA</td>
<td>b2317</td>
<td>conserved inner membrane protein</td>
<td>1b</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>S</td>
</tr>
<tr>
<td>EC958_2653</td>
<td>truA</td>
<td>b2318</td>
<td>pseudouridylate synthase I</td>
<td>0b</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>J</td>
</tr>
<tr>
<td>EC958_2822</td>
<td>bamB</td>
<td>b2512</td>
<td>lipoprotein required for OM biogenesis, in BamABCDEF complex</td>
<td>3</td>
<td>WT</td>
<td>0.063</td>
<td>WT</td>
<td>S</td>
</tr>
<tr>
<td>EC958_3571</td>
<td>nusA</td>
<td>b3169</td>
<td>transcription termination/antitermination L factor</td>
<td>0b</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>K</td>
</tr>
<tr>
<td>EC958_3581</td>
<td>greA</td>
<td>b3181</td>
<td>transcript cleavage factor</td>
<td>2</td>
<td>WT</td>
<td>0.063</td>
<td>WT</td>
<td>K</td>
</tr>
<tr>
<td>EC958_3621</td>
<td>sspA</td>
<td>b3229</td>
<td>stringent starvation protein A</td>
<td>2</td>
<td>WT</td>
<td>0.063</td>
<td>WT</td>
<td>O</td>
</tr>
<tr>
<td>EC958_4029</td>
<td>waaL</td>
<td>b3622</td>
<td>O-antigen ligase</td>
<td>3</td>
<td>LA</td>
<td>0.063</td>
<td>WT</td>
<td>M</td>
</tr>
<tr>
<td>EC958_4030</td>
<td>waaK</td>
<td>b3623</td>
<td>Lipopolysaccharide core biosynthesis; heptosyl transferase IV</td>
<td>3</td>
<td>LA</td>
<td>0.063</td>
<td>WT</td>
<td>M</td>
</tr>
<tr>
<td>EC958_4032</td>
<td>waaY</td>
<td>b3625</td>
<td>Lipopolysaccharide core heptose (II) kinase</td>
<td>5</td>
<td>LA</td>
<td>WT</td>
<td>&lt;0.3</td>
<td>T</td>
</tr>
<tr>
<td>EC958_4033</td>
<td>waaI</td>
<td>b3626</td>
<td>UDP-glucose:glucosyl</td>
<td>LPs</td>
<td>1,2-glucosyltransferase</td>
<td>2</td>
<td>WT</td>
<td>0.063</td>
</tr>
<tr>
<td>EC958_4034</td>
<td>waaO</td>
<td>b3627</td>
<td>UDP-Glc:Glucosyl</td>
<td>LPs</td>
<td>1,3-glucosyltransferase</td>
<td>5</td>
<td>X</td>
<td>0.063</td>
</tr>
<tr>
<td>EC958_4035</td>
<td>waaP</td>
<td>b3628</td>
<td>UDP-Gal:Glucosyl</td>
<td>LPs</td>
<td>1,6-D-galactosyltransferase</td>
<td>4</td>
<td>X</td>
<td>WT</td>
</tr>
<tr>
<td>EC958_4037</td>
<td>waaQ</td>
<td>b3630</td>
<td>Lipopolysaccharide core heptose (II) kinase</td>
<td>6</td>
<td>X</td>
<td>0.016</td>
<td>WT</td>
<td>-</td>
</tr>
<tr>
<td>EC958_4038</td>
<td>waaR</td>
<td>b3631</td>
<td>Lipopolysaccharide glucosyltransferase</td>
<td>6</td>
<td>X</td>
<td>0.016</td>
<td>0.5</td>
<td>M</td>
</tr>
<tr>
<td>EC958_4050</td>
<td>pyeA</td>
<td>b3642</td>
<td>orotate phosphoribosyltransferase</td>
<td>2</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>F</td>
</tr>
<tr>
<td>EC958_4246</td>
<td>wecA</td>
<td>b3784</td>
<td>undecaprenyl-phosphate alpha-N-acetylglucosamine transferase</td>
<td>5</td>
<td>LA</td>
<td>0.063</td>
<td>WT</td>
<td>M</td>
</tr>
<tr>
<td>EC958_4247</td>
<td>wzeE</td>
<td>b3785</td>
<td>ECA polysaccharide chain length modulation protein</td>
<td>2</td>
<td>X</td>
<td>0.063</td>
<td>WT</td>
<td>M</td>
</tr>
<tr>
<td>EC958_4252</td>
<td>wecD</td>
<td>b3790</td>
<td>TDP-fucosamine acetyltransferase</td>
<td>5</td>
<td>X</td>
<td>0.063</td>
<td>WT</td>
<td>-</td>
</tr>
<tr>
<td>EC958_4253</td>
<td>wecE</td>
<td>b3791</td>
<td>TDP-4-oxo-6-deoxy-D-glucose transaminase</td>
<td>2</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>M</td>
</tr>
</tbody>
</table>
investigated. Since the O25 antigen structure is known, we attempted to predict the glycosidic link formed by each GT using a combination of sequence similarity and O-antigen structure comparison as previously described [64]; the results of which are shown in Table 3 and Figure 5B. We were only able to mutate EC958_2371, while EC958_2372 and EC958_2375 were defined as essential in this study. Analysis of our EC958_2371 mutant confirmed that EC958_2371 is required for O-antigen biosynthesis (Figure S3), serum resistance and resistance to SDS (Table 2).

Other serum resistance mechanisms affecting LPS in EC958

LPS gel analysis was performed on all 54 defined mutants to identify genes that contribute to serum resistance by affecting LPS (Figure S3). The normal LPS pattern of EC958 consists of 12 bands including a thick bottom band representing the lipid A-core and an 11-band laddering pattern of lipid A-core bound O-antigen polymers, followed by approximately 6 thick bands of very long O-antigen chain length (Figure 5C). In addition to the 6 genes involved in O-antigen biosynthesis mentioned above, the LPS patterns of 20 mutants were altered in comparison to wild-type EC958; 6 of these mutants (waaL,K,J,F,G,I) only produced a lipid A-core (Table 2 and Figure S3).

LPS core biosynthesis genes. The genes involved in biosynthesis of the LPS outer core in EC958 share strong similarity with those from K-12 MG1655 (waaL and waaGPBIJYK) [reviewed in 65,66], suggesting that the LPS outer core structure in EC958 is the same as that in MG1655. As expected from the functions of waaGPBIJYK in MG1655, the waaG mutant showed the smallest size of lipid A-core, indicating that its whole outer core was not linked to the lipid A-inner core. Mutations in waaGPBIJYK also produced an expected LPS pattern of lipid A-core only, as these mutations prevent the linkage of O-antigen to the outer core. The LPS of the waaB and waaF mutants produced O-antigen laddering patterns containing an abnormal lipidA-core band, indicating that although the core structures were changed, these changes still allow the linking of O-antigen to the outer core. WaaL is responsible for the ligation of O-antigen to the outer core and mutation of this gene resulted in the synthesis of a lipid A-core without O-antigen. The EC958 waaL and waaGPBIJYK mutants were sensitive to human serum and SDS (except for the waaB and waaF mutant), while only the waaG, waaL and waaF mutants displayed enhanced sensitivity to NaCl (Table 2).

ECA biosynthesis genes. Seven serum resistant candidate genes within the ECA biosynthesis wec operon were identified by TraDIS (Table 2). Two of these genes (wzyE and wzyF) were not confirmed by serum assays to confer resistance and no changes were observed in the LPS patterns of these two mutants (Table 2, Figure S3). The wzyE mutant (but not the wzyF mutant) was, however, more sensitive to SDS (Table 2). The LPS patterns of the wzyE mutant revealed changes in the lipid A-core and in the modulation of O-antigen chain length (Figure S3). WzyE has previously been shown to preferentially form a protein complex with WzyE and WzzE for the biosynthesis of ECA over Wzy and Wzz [67]. However, our results suggested that WzyE might also contribute to the chain length regulation of O-antigen in EC958. WecA has been shown previously to be involved in the biosynthesis of O7, O18, O75, and O111 antigen [68] and our results indicate that the same is also true for O25b, thus explaining the role of WecA in serum resistance in EC958. The wecD mutant was highly sensitive to serum, possessed an LPS profile that was altered in both the lipid A-core and the amount of O-antigen, and displayed enhanced sensitivity to SDS (Table 2, Figure S3). Further investigation is required to understand the involvement of WecD in lipid A-core and O-antigen biosynthesis. Finally, none of the ECA biosynthesis mutants were altered in sensitivity to NaCl.

Other genes. The remaining 8 mutants with altered LPS patterns were all confirmed as serum sensitive (Table 2). They included one mutant, yfaH, which produced only lipid A-core, and 7 mutants with different O-antigen patterns: hyd (EC958_0460), hydR (EC958_0461), nagA (N-acetylglucosamine-6-phosphate deacetylase), pgm (phosphoglucomutase), galE (UDP-galactose-4-epimerase), EC958_1112 and wcaF (predicted acetyltransferase involved in colanic acid synthesis [69]). Some of these genes were also associated with resistance to additional stresses; EC958 yfaH and wcaF mutants were sensitive to SDS and NaCl, while nagA and pgm mutants were sensitive to SDS (Table 2).

Novel O-antigen chain length regulators

Mutation of the hyd and hydR genes in EC958 resulted in the modulation of O-antigen chain length (Figure 6B). The EC958

Table 2. Cont.

<table>
<thead>
<tr>
<th>EC958 locus_tag</th>
<th>Gene Name</th>
<th>Matched in MG1655</th>
<th>Product</th>
<th>Serum*</th>
<th>LPS†</th>
<th>SDS MIC‡ (%)</th>
<th>NaCl MIC‡ (M)</th>
<th>COG code</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC958_4254</td>
<td>wzyE</td>
<td></td>
<td>lipid III flippase</td>
<td>0b</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>R</td>
</tr>
<tr>
<td>EC958_4255</td>
<td>wecf</td>
<td>b4481</td>
<td>TDP-Fuc4NAcA-lipidFuc4NAc transferase</td>
<td>1b</td>
<td>WT</td>
<td>0.063</td>
<td>WT</td>
<td>-</td>
</tr>
<tr>
<td>EC958_4256</td>
<td>wzyE</td>
<td>b3793</td>
<td>ECA polysaccharide chain elongation</td>
<td>0b</td>
<td>WT</td>
<td>0.063</td>
<td>WT</td>
<td>-</td>
</tr>
<tr>
<td>EC958_4322</td>
<td>rfaH</td>
<td>b3842</td>
<td>DNA-binding transcriptional antitranslator</td>
<td>5</td>
<td>LA</td>
<td>0.016 &lt;0.3</td>
<td>K</td>
<td></td>
</tr>
<tr>
<td>EC958_4336</td>
<td>dsbA</td>
<td>b3860</td>
<td>Protein disulfide oxidoreductase</td>
<td>1</td>
<td>WT</td>
<td>0.031 0.6</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td>EC958_4480</td>
<td>pgi</td>
<td>b4025</td>
<td>glucosylphosphate isomerase</td>
<td>1b</td>
<td>WT</td>
<td>0.063</td>
<td>WT</td>
<td>G</td>
</tr>
<tr>
<td>EC958_4658</td>
<td>misA</td>
<td>b4171</td>
<td>Δ(2)-isopentenylprenylphosphate RNA-adenosine transferase</td>
<td>0b</td>
<td>WT</td>
<td>0.063</td>
<td>WT</td>
<td>J</td>
</tr>
<tr>
<td>EC958_4725</td>
<td>fbp</td>
<td>b4232</td>
<td>fructose-1,6-bisphosphatase I</td>
<td>3</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>G</td>
</tr>
</tbody>
</table>

*Serum sensitivity was shown as -log_{10} of viable count difference between t=90 min and t=0 min. The higher the number the more sensitive the mutant is.
†LPS patterns were coded as followed: WT, similar to wild-type strain, no change observed; LA, only the band for lipid A-core observed; LA+1, lipid A-core band and only one band above it; X, changes observed, see Supplementary Figure S3 for band patterns.
‡The wild-type (WT) MIC is 0.8 M for NaCl and 0.125% for SDS.

doi:10.1371/journal.pgen.1003834.t002
hyxA mutant exhibited an increased proportion of O-antigen chain of 2 to 6 units with maximum number at 3–5 units and reduction in very high chain length polymer. The EC958 hyxR mutant had an increased proportion of O-antigen polymer of 2–4 units. The hyxA and hyxR genes are located in a pathogenicity island (PAI-X) consisting of 4 genes (fimX and hyxRAB) as previously described in the UPEC strain UTI89 [70]. The hyxB gene (EC958_0459) has also been named upaB due to its function as an autotransporter [71], and we prefer to maintain this nomenclature. This island is present (in several variations) in 24 out of 50 E. coli genomes across all pathotypes (Figure 6A) and current sequence data suggest that it is exclusive to E. coli.

No functional prediction was found for hyxA, and thus this is the first report of hyxA involvement in serum resistance by regulating O-antigen chain length. The hyxR gene encodes a LuxR-like response regulator that suppresses the nitrosative stress response and contributes to intracellular survival in macrophages by regulating hmpA, which encodes a nitric oxide-detoxifying flavohaemoglobin [70]. The expression of the hyxR gene is regulated through bidirectional phase inversion of its promoter region by the upstream gene fimX, which encodes a tyrosine-like recombinase [70]. It is also worth noting that the contribution of hyxA to serum resistance was greater than hyxR, as demonstrated by the 3-log reduction in viability by the hyxA mutant compared to the hyxR mutant. A serum sensitive phenotype for the hyxR mutant was only observed in mixed competition assays, and both mutants did not exhibit altered sensitivity to SDS or NaCl (Table 2).

**Novel serum resistance mechanisms not affecting LPS**

A major advantage of whole genome approaches such as TraDIS lies in their power of discovery. Out of 46 genes that define the serum resistome of EC958, 21 (46%) genes were confirmed to be required for serum resistance independent of altered LPS patterns (Table 2; Figure S3). The function of these genes ranged across 10 COG functional categories and included ‘Carbohydrate transport and metabolism’ (4 genes), ‘Cell wall/membrane/envelope biogenesis’ (3), ‘Posttranslational modification, protein turnover, chaperones’ (3) and many others (Table 2). Twelve of these genes were also associated with enhanced sensitivity to SDS and one with enhanced NaCl sensitivity (Table 2).

Of the non-LPS genes required for serum resistance, the most notable was lpp (logFC = 10, log difference = 6) (Table 2). The lpp gene encodes one of the most abundant proteins in E. coli and is responsible for the stabilisation and integrity of the bacterial cell envelope [72]. Mutation of the lpp gene results in the formation of outer membrane blebs, leakage of periplasmic enzyme ribonuclease, decreased growth rate in media of low ionic strength or low osmolarity and hypersensitivity to toxic compounds [73,74]. Indeed, the EC958 lpp mutant was more sensitive to SDS, suggesting decreased membrane integrity (Table 2). To the best of our knowledge, this study is the first to show a direct link between Lpp and serum resistance.

Another set of genes notable in our TraDIS analysis include tolQAB, which encode three of the six proteins (YbgC-YbgF-TolQ-R-A-B-Pal) that make up the Tol-Pal system of E. coli cell
envelope. The Tol-Pal system is responsible for maintaining the integrity of the outer membrane. TolQRA form an inner membrane complex in which TolQR is necessary for its stability [75]. TolB, a periplasmic protein, connects the inner membrane complex with the peptidoglycan-associated lipoprotein, Pal, which is anchored to the outer membrane [76]. TolB, a periplasmic protein, connects the inner membrane complex in which TolQR is necessary for its stability [75].

BamB is a lipoprotein that is part of the BamABC complex. Mutation in _bamB_ results in increased outer membrane permeability, thus enhancing sensitivity to rifampin and dramatically reducing growth on SDS and novobiocin [77]. Our data showed that mutation of the _bamB_ gene in EC958 resulted in increased sensitivity to both human serum and SDS (Table 2).

Our TraDIS experiment also indicated that the modification of lipid A with L-Ara4N was important for serum resistance. Three (arnDEF) of the seven genes involved in the biosynthesis and attachment of L-Ara4N to lipid A-core were identified as part of the serum resistome of EC958 and their role was confirmed by mutagenesis (Table 2). This mechanism is known to confer resistance to polymixin B by preventing its binding to lipid A [Reviewed in 78,79,80]. ArnD catalyzes a deformylation step to generate UDP-L-Ara4N before it is transported across the inner membrane by ArnEF [79,81]. The requirement of ArnDEF for serum resistance indicates that EC958 requires L-Ara4N modification to evade the antimicrobial activity of cationic peptides present in human serum. Interestingly, only the _arnD_ mutation conferred sensitivity to SDS (Table 2), which might suggest a role of UDP-L-Ara4N in maintaining membrane integrity. Further investigation is needed to understand why ArnT, the final enzyme of UDP-L-Ara4N in maintaining membrane integrity. Further investigation is needed to understand why ArnT, the final enzyme

<table>
<thead>
<tr>
<th>Locus tag</th>
<th>Gene</th>
<th>No. of amino acids</th>
<th>Similar protein</th>
<th>% identities/% positives/no. of amino acid overlap</th>
<th>predicted function</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC958_2381</td>
<td>rmb</td>
<td>361</td>
<td>dTDP-glucose 4,6-dehydratase [Escherichia coli NA114]</td>
<td>99/100/361</td>
<td>dTDP-glucose 4,6-dehydratase</td>
</tr>
<tr>
<td>EC958_2380</td>
<td>rmd</td>
<td>299</td>
<td>dTDP-4-dehydroxyarabonose reductase [Escherichia coli NA114]</td>
<td>100/100/299</td>
<td>dTDP-4-dehydroxyarabonose reductase</td>
</tr>
<tr>
<td>EC958_2379</td>
<td>rmr</td>
<td>292</td>
<td>glucose-1-phosphate thymidylytransferase (Escherichia coli NA114)</td>
<td>100/100/292</td>
<td>glucose-1-phosphate thymidylytransferase</td>
</tr>
<tr>
<td>EC958_2378</td>
<td>rmc</td>
<td>180</td>
<td>dTDP-4-dehydroxyarabonone 3,5-epimerase (Escherichia coli NA114)</td>
<td>99/99/180</td>
<td>dTDP-4-dehydroxyarabonone 3,5-epimerase</td>
</tr>
<tr>
<td>EC958_2377</td>
<td>wxx</td>
<td>419</td>
<td>O-antigen flippase Wxx [Escherichia coli E47a] (AD43260)</td>
<td>95/98/149</td>
<td>O-antigen flippase</td>
</tr>
<tr>
<td>EC958_2376</td>
<td>wekA</td>
<td>321</td>
<td>putative glycosyl transferase family 2 [Escherichia coli ABU83972] (ADN46883)</td>
<td>93/98/321</td>
<td>glycosyl transferase family 2; α-L-FucpNAc-(1-&gt;3)-β-D-GlcNAc</td>
</tr>
<tr>
<td>EC958_2375</td>
<td>wekB</td>
<td>382</td>
<td>putative glycosyl transferase [Escherichia coli E47a] (AD43262)</td>
<td>93/97/382</td>
<td>glycosyl transferase family 4; β-D-Glc-(1-&gt;6)-α-D-Glc</td>
</tr>
<tr>
<td>EC958_2374</td>
<td>wzy</td>
<td>405</td>
<td>O-antigen polymerase Wzy [Escherichia coli E47a] (AD43263)</td>
<td>92/95/346</td>
<td>O-antigen polymerase</td>
</tr>
<tr>
<td>EC958_2373</td>
<td>wbjJ</td>
<td>201</td>
<td>predicted lipopolysaccharide biosynthesis</td>
<td>68/80/194</td>
<td>predicted lipopolysaccharide biosynthesis O-acetyl transferase</td>
</tr>
<tr>
<td>EC958_2372</td>
<td>wbuH</td>
<td>366</td>
<td>Putative glycosyltransferase WbuH [Escherichia coli O4K3H35] (AAAT8564)</td>
<td>29/49/372</td>
<td>glycosyl transferase family 4; α-D-Glc-(1-&gt;3)-α-L-FucpNAc</td>
</tr>
<tr>
<td>EC958_2371</td>
<td>wblL</td>
<td>263</td>
<td>Rhamnosyltransferase [Escherichia coli str. K-12 substr. MG1655] (P36687)</td>
<td>60/76/260</td>
<td>Rhamnosyltransferaseα-L-Rhap(1-&gt;3)α-D-Glc</td>
</tr>
<tr>
<td>EC958_2370</td>
<td>gnd</td>
<td>468</td>
<td>6-phosphogluconate dehydrogenase [Escherichia coli NA114] (YP_006137974.1)</td>
<td>100/100/468</td>
<td>6-phosphogluconate dehydrogenase</td>
</tr>
<tr>
<td>EC958_2369</td>
<td>ugd</td>
<td>388</td>
<td>UDP-glucose-6-dehydrogenase [Escherichia coli NA114] (YP_006137975)</td>
<td>100/100/388</td>
<td>UDP-glucose-6-dehydrogenase</td>
</tr>
<tr>
<td>EC958_2368</td>
<td>wzz</td>
<td>325</td>
<td>regulator of length of O-antigen component of lipopolysaccharide chain [Escherichia coli NA114] (YP_006137976)</td>
<td>100/100/325</td>
<td>O-antigen chain length regulator</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pgen.1003834.t003
GDP-mannose or GDP-glucose in the cell and thus regulating cell wall biosynthesis [82]. Both Pgi and Fbp catalyze the production of D-fructose-6-phosphate from β-D-glucose-6-phosphate and fructose-1,6-bisphosphate, respectively [83,84]. D-fructose-6-phosphate is a precursor for the biosynthesis of UDP-GlcNAc, which in turn is required for peptidoglycan, lipid A and ECA biosynthesis. Thus, these three enzymes may catalyze key reactions that, if disrupted, could adversely affect the downstream biosynthesis of cell surface components including colanic acid, peptidoglycan, lipid A and ECA.

An ompA mutant is protected from serum killing when present at a low proportion in a mixed bacterial population

Of all the chromosomal genes previously attributed to serum resistance, the only gene that was not identified in our TraDIS screen was ompA. To examine this further we constructed an EC958 ompA mutant and indeed observed it was sensitive to killing by human serum (Figure 7). One way to explain this discrepancy is that the phenotype of an ompA mutant could be complemented in trans by other ompA-intact bacteria in a mixed population such as the mutant library. In fact, OmpA inhibits serum-mediated killing by binding to C4b-binding protein (C4BP) to prevent the activation of C3b via the classical complement pathway [85], and OmpA is known to be released from E. coli cells when treated with serum [86]. In our mutant library, ompA mutants only accounted for approximately 0.02% of the total bacterial cells, and thus we hypothesized that the release of OmpA from 99.98% of the cells, when treated with serum, provided OmpA in trans to complement ompA mutants. We tested this hypothesis by mixing the ompA mutant with wild-type EC958 at various ratios and indeed showed that ompA mutants were protected from serum killing if the proportion of ompA mutants was less than 15% (Figure 7). This result strongly suggests that in trans complementation of OmpA prevents the identification of ompA as a serum resistance gene in our assay.

Complementation of selected mutants restores serum resistance

To further demonstrate the function of non-LPS genes in serum resistance, we selected three orphan genes (acnB, greA and fbp) that do not belong to an operon to perform genetic complementation.

---

**Figure 6. Genetic context of the hyxA and hyxR genes and their role in O-antigen synthesis.** (A) The conserved location of PAI-X containing hyxA and hyxR genes in E. coli genomes of various pathotypes. (B) The LPS pattern of the hyxA and hyxR mutants and their complemented strains, demonstrating a role for the hyxA and hyxR genes in O-antigen chain length regulation. Red boxes highlight differences in LPS patterns compared with that of the wild-type.

doi:10.1371/journal.pgen.1003834.g006
For these experiments, the selected gene was amplified by PCR, cloned into the low copy number plasmid pSU2718G, and transformed into the respective mutant strain for complementation. In each case, the phenotype of the complemented strain exactly matched that of wild-type EC958 (Table 4). Taken together, these results confirm the role of acnB, gela1 and fhp in serum resistance and provide further evidence to support the use of techniques such as TraDIS in functional gene discovery.

Discussion

The rapid advancement of new sequencing technologies has created novel opportunities to interrogate biological systems that were not previously possible. TraDIS was first described as a method that combined high-density mutagenesis with Illumina next generation sequencing technology to study the essential genes of S. Typhi and the conditional essential genes required for survival in bile [42]. The increased data output afforded by next generation DNA sequencing is particularly useful and cost-effective for small bacterial genomes, however it presents technical and bioinformatical challenges for applications such as TraDIS that utilize low complexity DNA libraries. Here we present the application of a modified version of TraDIS that is amenable to multiplexing using the Illumina HiSeq 2000 platform, and we demonstrate its effectiveness by using it to define the essential gene repertoire and the serum resistome of a multidrug resistant strain from the globally disseminated E. coli ST131 lineage.

The multiplexed TraDIS protocol utilizes a newly designed custom oligonucleotide in the library enrichment step of the Illumina library preparation protocol (Figure S1). This oligonucleotide incorporates the Illumina sequencing primer-binding site into transposon specific DNA fragments, enabling the use of the standard Illumina sequencing primer and eliminating the need to design and optimize another sequencing primer for each new transposon sequence. The 6-bp barcode immediately after the sequencing primer-binding site allows 12-sample multiplexing within one lane. The use of 12 barcodes at the first 6 nucleotides of read 1 increased the complexity of the library compared with the original TraDIS protocol, thus reducing data loss due to mis-identification of clusters [87,88]. However, the number of useable reads from our sequencing runs was still low (15–20% of total reads). We believe non-specific amplification in the enrichment step was the main cause, and further optimization of the enrichment PCR conditions is required. Similar approaches combining transposon mutagenesis with high-throughput sequencing (Tn-seq [45], INSeq [43], HITS [44]) have also been used to address different scientific questions, including the identification of essential genes and genes associated with enhanced fitness in specific growth conditions [41,42,45], determination of niche-specific essential genes [43,44,89], identification of genes associated with tolerance to various agents/conditions [42,90] and many other applications as reviewed elsewhere [91,92]. In terms of insertion density, we achieved 502,068 independent insertion sites with a density of 1/10 bp (i.e. one insertion every 10 bp), which is comparable with the work by Christen et al. (1/8 bp in C. crescentus) [41], Langridge et al. (1/13 bp in Salmonella Typhi) [42] and Barquist et al. (1/9 bp in Salmonella Typhimurium) [46].

The identification of the essential gene set for a single organism is challenging due to several factors, including the presence of transposon insertion cold spots (i.e. regions of low transposon insertion frequency), the difficulty in distinguishing mutations that prevent growth from those that severely reduce growth rate, pre-existing gene duplications and the specific growth conditions used in the experiment [93–95]. In this study, we define essential genes as those that, when mutated by transposon insertion, either prevent or severely attenuate growth on LB agar media supplemented with 30 μg/ml Cm at 37°C. The cut-off value to determine whether a gene is essential was defined as the intercept of two distributions of the insertion index of each gene: the exponential distribution representing essential genes and the gamma distribution representing non-essential genes [42]. This means that our essential genes also include those genes that can tolerate insertions but were severely attenuated in the input pool. Out of 315 essential genes, 64 genes had no transposon insertions, 178 genes had 1 to 5 transposon insertions and 73 genes had more than 5 transposon insertions (Table S1).

The high density of insertion sites achieved in our study provided reliable data for the identification of essential genes within the EC958 genomes with a minimal probability of false positive calls due to transposon insertion cold spots. The identification of essential genes has previously been performed using several approaches in E. coli K-12 (strains MG1655 and W3110) [96,97]. Baba et al. generated null mutations by lambda-mediated recombination in 3905 E. coli W3110 genes (the Keio library), but were unable to mutate 303 candidate essential genes [97]. This set of essential genes was further consolidated by manual literature review on the EcoGene website (www.ecogene.org), which reduced the set to 289 genes. Of the 315 essential genes identified for EC958 in this study, 291 genes (73%) matched those previously described in the EcoGene list (Table S1). There were 84 essential genes specific for EC958, twenty-four of which do not have homologs in the MG1655 genome. In contrast, 58 genes previously identified as essential for E. coli K-12 were either not present in EC958 or not identified in our analysis. The majority of essential genes in EC958 are conserved with 91% of the genes present in more than 90% of E. coli complete genomes available.

In this study, we provided two layers of evidence for the role of each serum resistance gene: by simultaneously assaying a large mutant library and by generation of defined mutants for independent phenotypic testing. Indeed, using defined mutagenesis we were able to confirm a role for 46 of the 56 genes identified as essential genes for serum resistance.

Figure 7. In trans complementation of wild-type EC958 protects ompA mutants from serum killing. Wild-type EC958 and an EC958 ompA mutant were mixed at different ratios, exposed to human serum for 90 minutes, and the EC958 ompA mutant was examined for its survival in the mixed bacterial suspension.
by TraDIS in serum resistance. To the best of our knowledge, this represents the first large scale follow-up of TraDIS data in this manner and highlights the effectiveness of the technique in large-scale functional genomics. Our study also revealed that trans complementation of specific mutants can occur in a large mutant library population, as demonstrated by our findings with an \textit{acnB} mutant. We also demonstrated complete complementation of mutants containing deletions in the \textit{acoB}, \textit{greA} and \textit{fhp} genes, corroborating their novel role in serum resistance independent of LPS alterations. Finally, we provided further insight into the mechanistic action of the serum resistance genes identified in EC958 by examining the survival of the respective mutants to outer membrane stresses that affect antimicrobial access and osmotic potential.

The search for genetic determinants of serum resistance in bacteria has been ongoing since the 1970s [98]. Our current understanding of the mechanisms that promote bacterial resistance to human serum include a role for surface structures such as O antigens, K antigens, outer membrane proteins (OmpA), and plasmid-encoded proteins (TraT, Iss) [37–39]; notably, however, not all of these mechanisms are required for resistance in a single bacterial strain [32,33]. Our study represents the first report to simultaneously investigate the entire serum resistome of one strain. Our results demonstrated that both the lipid A-core and O25 antigen are crucial for serum resistance in EC958, while K antigen does not contribute to serum resistance. Out of 54 defined mutants investigated, half had changes in their LPS gel patterns, all of which resulted in serum resistance. In contrast, none of the K capsular biosynthesis genes were identified in our TraDIS screen. This result is similar to that reported for the O75::K5 UPEC strain GR-12, where alterations in O75 LPS affected serum resistance more than a K5 null mutation [34]. It is likely, however, that there are strain-specific differences for the role of O antigen and K capsule in serum resistance, and that this reflects differences in the make-up of these structures. For example, previous analysis of an \textit{E. coli} O4:K5:H5 blood isolate revealed that the K54 antigen contributes more to serum resistance than the O4 antigen [99]. Other K antigens such as the K1 and K2 capsules have also been shown to play an important role in serum resistance [100,101]. The K antigen expressed by EC958 has not been typed but genomic analysis shows that EC958 has a group 2 capsular gene cluster that conforms to the conserved structure of this group [Reviewed in 102]. However, region 2, which encodes glycosyltransferases specific for each K type, shares such low similarity with available sequences in the GenBank database that deducing its K type \textit{in silico} was not possible.

The O25 antigen gene cluster was further characterized using sequence analysis, targeted mutation and LPS profiling. All of the \textit{dtDP-\alpha-L-rhamnose biosynthesis genes (\textit{m}i\textit{CADB}) were required for serum resistance. However, EC958 lacks the biosynthesis genes for UDP-Fuc-Nac, a component of O25 antigen unit. If, based on the cross reaction of antiserum against the O25 antigen with O25b expressing cells, we assume that EC958 has the same O-antigen repeat unit as the O25 determined from previous studies [57,58], then EC958 must possess a novel mechanism for the synthesis or uptake of UDP-Fuc-Nac.

Two additional surface antigens that contribute to serum resistance in EC958 are the enterobacterial common antigen and colanic acid (M antigen). Mutations in five ECA biosynthesis genes rendered EC958 susceptible to serum. While mutation of three of these genes (\textit{wccA}, \textit{wzzE} and \textit{wccD}) affected LPS and sensitivity to SDS, mutation of \textit{wccE} and \textit{wecF} did not change LPS (although a \textit{wecF} mutant was more sensitive to SDS), suggesting that the ECA may be involved in serum resistance, perhaps indirectly via its role in membrane integrity. Our data also suggest the involvement of colanic acid in serum resistance. Three genes encoding for colanic acid biosynthesis (\textit{wecI}, \textit{gmm} and \textit{wecF}) were identified by TraDIS. Gmm is most likely to be involved in the biosynthesis of colanic acid [82], while mutation of the \textit{wccI} gene did not confer serum resistance. The product of \textit{wccI} was predicted to be an acetyltransferase [103] required for colanic acid production [69]. The \textit{wccI} \textit{wecF} mutant possessed an altered LPS pattern with a reduced amount of O-antigen (especially very long chain length O-antigen) and was sensitive to both SDS and high osmolality. The enhanced sensitivity of the EC958 \textit{wccI} \textit{wecF} mutant could therefore be explained by a number of factors, including altered LPS, altered colanic acid and reduced overall membrane integrity.

A number of other genes were identified that contributed to serum resistance in an LPS-dependent manner. The gene \textit{nagA} encodes N-acetylglucosamine-6-phosphate deacetylase, an enzyme important for the metabolism of N-acetyl-D-glucosamine [104]. It catalyzes the first step in producing UDP-GlcNac, a nucleotide sugar required for ECA, lipid A and peptidoglycan biosynthesis, by deacetylating N-acetylglucosamine-6-phosphate to glucosamine-6-phosphate [Reviewed in 105]. However, \textit{NagA} is not solely responsible for the production of UDP-GlcNac because glucosamine-6-phosphate can also be obtained via GlmS from fructose-6-phosphate or taken up from the environment by \textit{ManXYZ} [105]. Indeed, the LPS banding pattern of the \textit{nagA} mutant was different to that of the parent strain (i.e. it possessed thicker second and third bands from the bottom of the gel; Figure S3), suggesting its enhanced sensitivity phenotype may be associated with a predominantly shorter O antigen.

\textit{Pgm} is a phosphoglucomutase that catalyses the reversible conversion of glucose-1-phosphate to glucose-6-phosphate, an important step in galactose and maltose catabolism [106]. A \textit{pgm} mutant has several phenotypes; it is defective in swimming and swarming mobility [107], it possesses an aberrant (shorter and wider) cell morphology, is sensitive to detergents [108] and it stains blue with iodine when grown in the presence of galactose [109]. An \textit{EC958 pgm} mutant produced little full length O-antigen; the majority of its LPS condensed into a thick band of incomplete lipid A-core and a thin clear band of lipid A-core plus one unit of O-antigen. This feature is consistent with the high serum and SDS sensitivity phenotypes observed for this mutant.

\textit{GalE} is a well-studied enzyme that catalyzes the interconversion of UDP-galactose and UDP-glucose [Reviewed in 109]. Both nucleotide sugars are required for colanic acid biosynthesis.

---

**Table 4. Complementation of selected mutants.**

<table>
<thead>
<tr>
<th>Strains</th>
<th>Phenotype</th>
<th>Serum</th>
<th>SDS (%)</th>
<th>NaCl (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC958 (wild-type)</td>
<td>0</td>
<td>0.125</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>EC958\textit{AcnB}</td>
<td>1</td>
<td>0.125</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>EC958\textit{AcnB} + pAcnB</td>
<td>0</td>
<td>0.125</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>EC958\textit{GreA}</td>
<td>2</td>
<td>0.063</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>EC958\textit{GreA} + pGreA</td>
<td>0</td>
<td>0.125</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>EC958\textit{fhp}</td>
<td>3</td>
<td>0.125</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>EC958\textit{fhp} + pFhp</td>
<td>0</td>
<td>0.125</td>
<td>0.8</td>
<td></td>
</tr>
</tbody>
</table>

*Serum sensitivity was shown as \( \log_{10} \) of viable count difference between \( t = 90 \) min and \( t = 0 \) min. The higher the number the more sensitive the mutant is.

**Minimal inhibitory concentration (MIC) of SDS (%) and NaCl (M).**

doi:10.1371/journal.pgen.1003834.t004

---
Furthermore, UDP-glucose is used in three steps to synthesize the LPS outer core (catalyzes by WaaG, WaaL and WaaJ). LPS patterns of the galE mutant exhibited a very thick band of lipid A-core, suggesting that the lipid A-core in this strain has multiple sizes. This may be explained by the limiting effect of UDP-glucose in the three steps involved in its incorporation into the outer core. UDP-glucose can also be synthesized by GaldU from glucose-1-phosphate [110], which may explain why an EC958 galE mutant could still make LPS (Figure S3). Despite being able to make LPS, however, the galE mutant was sensitive to human serum. Whether this sensitivity can be attributed to the effect a galE mutation has on LPS or colanic acid remains to be determined.

We have demonstrated the successful application of multiplexed TraDIS for a functional genomics study targeted at E. coli EC958, a prototype strain from the globally disseminated and multidrug resistant E. coli ST131 lineage. This approach enabled the first description of an essential gene set from an ExPEC strain. Our work has also defined the serum resistome in E. coli EC958. This comprehensive inventory of E. coli EC958 genes that contribute to this phenotype provides a framework for the future characterization of virulence genes in ExPEC as well as other Gram-negative pathogens that cause systemic infection.

Materials and Methods

Ethics statement

Approval for the collection of human blood was obtained from the University of Queensland Medical Research Ethics Committee (2008001123). All subjects provided written informed consent.

Bacterial strains and growth conditions

E. coli EC958 was isolated from the urine of a patient presenting with community UTI in the Northwest region of England and is a representative member of the UK epidemic strain A (PFGE type), one of the major pathogenic lineages causing UTI across the United Kingdom [23]. EC958 Δlac, which contained a mutation in the lac operon, was used in competitive assays. This strain had an identical growth rate to wild-type EC958. Strains were routinely supplemented with the appropriate antibiotics (Cm 30 μg/ml or gentamicin 20 μg/ml) unless indicated otherwise.

Generation of miniTn5-Cm mutant library

A miniTn5-Cm transposon containing a Cm cassette flanked by Tn5 mosaic ends (sequence from Epicenter) was PCR amplified from pKD3 plasmid DNA (Vald digested) using primers 2279 5’-CTGTCTCTTTACATACATCtcagttggtggtttggagc-3’ and 2280 5’-CTGTCTCTTTACATACATCGcagttggtggtttggagc-3’. The PCR reactions were performed using Phusion High-Fidelity DNA polymerase (New England BioLabs). The amplicon was purified using the QIAGEN MinElute PCR purification kit before being phosphorylated using T4 polynucleotide kinase (New England BioLabs) and subjected to the final purification step. A total of at least 800 ng of this miniTn5-Cm transposon DNA was incubated in an 8 μl reaction containing 4 μl of EZ-Tn5 transposase (Epicenter Biotechnologies) at 37°C for 1 h then stored at −20°C.

Bacterial cells were prepared for electroporation as previously described [42]. Briefly, cells were grown in 2×TY broth to an OD_{600} of 0.3-0.5, then harvested and washed three times in 0.5× volume of 10% cold glycerol before being resuspended in a 1/1000× volume of 10% cold glycerol and kept on ice. A volume of 60 μl cells was mixed with 0.2 μl of transposomes and electroporated in a 2 mm cuvette using a BioRad GenePulser set to 2.5 kV, 25 μF and 200Ω. Cells were resuspended in 1 mL SOC medium and incubated at 37°C for 2 hours, then spread on LB agar plates supplemented with Cm 30 μg/mL. After incubation overnight at 37°C, the total number of colonies was estimated by counting a proportion from multiple plates. Chloramphenicol resistant colonies were resuspended in sterilised LB broth using a bacteriological spreader before adding sterile glycerol to 15% total volume and stored in −80°C. Each batch of mutants contained an estimated 32,000 to 180,000 mutants. The final library of 1 million mutants was created by pooling 11 mutant batches, resulting in a cell suspension of 2×10^{11} CFU/ml.

Transposon library screening in human serum

Freshly pooled human serum was collected from at least two healthy individuals on the day of the experiment. Ten milliliters of blood was collected from each person and centrifuged at 4000 rpm for 10 minutes to collect the serum. Approximately 2×10^{10} viable mutants were incubated in 1 ml of 50% freshly pooled human serum in LB broth at 37°C for 90 minutes. The control samples were prepared in the same way but were incubated with inactivated serum (Millipore) instead of fresh serum. Both control and test experiments were performed in duplicate. The cells were then washed twice with sterile 1× PBS to remove serum, transferred to 100 ml LB broth and allowed to grow at 37°C with 250 rpm shaking for 4 hours. The genomic DNA was then extracted from 5 ml of each culture using Qiagen 100-G genomic tips.

Multiplexed TraDIS

Genomic DNA was standardized to 3.6 μg in a volume of 120 μl before being sheared by Covaris S2 according to the Illumina TruSeq Enrichment gel-free method (TruSeq DNA sample preparation v2 guide). The subsequent steps of DNA end repair, DNA end adenylation and adapter ligation were also done following the Illumina TruSeq v2 instructions. The adapter-ligated fragments containing transposon insertion sites were enriched using a transposon-specific indexing forward primer and the Illumina reverse primer Index 1 (Table S5) at 500 nM each per reaction. This 89 bp forward primer binds specifically to miniTn5-Cm transposon (25 bp) and carries 64 bp overhang which includes 6 bp index sequence, 33 bp binding site for Illumina read 1 sequencing primer and 23 bp P5 sequence for binding to the flowcell. The enrichment step was done using the KAPA Library Amplification kit (KAPA Biosystems) at an annealing temperature of 60°C for 22 cycles. The KAPA Library Quantification kit was used to measure the concentration of DNA fragments in the enriched library. Twelve libraries from 12 samples were pooled to equimolar concentrations to achieve a cluster density of 850 K/mm² when 10 nM of library pool was loaded onto the flowcell. The 12-plex pool was loaded on 3 lanes of TruSeq v2 and 3 lanes of TruSeq v3 flowcells for sequencing using a 100 cycles, paired-end protocol to access the reproducibility and read quantity among lanes and flowcells. The data from six samples (Figure 1A) were presented in this study. The TraDIS sequence data from this study was deposited on the Sequence Read Archive (SRA) under the BioProject number PRJNA189704.

Analysis of nucleotide sequence data

Sequence reads from the FASTQ files were split according to twelve 6 bp index sequences combined with the 37 bp transposon-specific sequence using fastx_barcode_splitter.pl (total length of 43 bp as barcodes, allowing for 2 mismatches) (FASTX-Toolkit version 0.0.13, http://hannonlab.cshl.edu/fastx_toolkit/index.html). The barcode matching reads were trimmed off the 43 bp
barcode at the 5’ end and 25 bp of potential low quality at the 3’ end, resulting in high quality sequence reads of 31 bp in length that were used to map to the EC958 chromosome (PRJEA61443) by Maq version 0.7.1 [111]. Subsequent analysis steps were carried out as previously described [42] to calculate the number of sequence reads (raw read counts) and the number of different insertion sites for every gene, which were then used to estimate the threshold to identify essential genes. The read counts and insertion sites were visualized using Artemis version 13.0 [112]. The circular genome diagram was generated by CGView [113] and linear genetic comparison was illustrated using Easyfig version 2.1 [114].

Statistical analyses

We identified genes required for survival in human serum by comparing the differences in read abundance of each gene between the inactivated serum control and active serum test samples using the Bioconductor package edgeR (version 2.6.10) [48]. The raw read counts from two biological replicates of each treatment were loaded into the edgeR package (version 2.6.12) using the R environment (version 2.15.1). Genes that have very low read counts in all the samples (essential genes) were removed from further analysis. The composition bias in each sequence library was normalized using the trimmed mean of M value (TMM) method [115]. We then used the quantile-adjusted conditional maximum likelihood (qCML) for negative binomial models to estimate the dispersions (biological variation between replicates) and to carry out the exact tests for determining genes with significantly lower read counts in the test samples compared to the control samples [116,117]. Stringent criteria of log fold-change ($\log_{2}FC \leq -1$) and false discovery rate $\leq 0.001$ were chosen to define a list of the most significant genes for further investigation by phenotypic assays.

Molecular methods

Chromosomal DNA purification, PCR and DNA sequencing of PCR products was performed as previously described [118]. Defined mutations were made using the $\lambda$-Red recombinase method with some modifications [22,47]. In brief, the final PCR products were fused and amplified from three fragments containing two 500-bp homologous regions flanking the gene of interest and a Cm cassette from pKD3 plasmid (see Table S5 for list of primers). The fused PCR products were then electroporated into EC958 harbouring a gentamicin resistant plasmid carrying the $\lambda$-Red recombinase gene. Mutants were then selected and confirmed by sequencing. Completion was done by cloning the gene of interest into a gentamicin resistant derivative of pSU2718 [119] at BamHI-XhoI cut sites (primers listed in Table S5). The construct was then transformed into the respective mutant and induced using 1 mM IPTG before and during phenotypic assays.

Serum resistance assay

Overnight bacterial cultures were washed in phosphate buffered saline (PBS) and then standardized to an OD$_{600}$ of 0.8. Equal volumes (50 $\mu$L) of standardized cultures and pooled human sera were mixed and incubated for 90 min at 37°C (in triplicates). Viable counts were performed to estimate the number of bacterial cells prior to serum treatment ($t=0$ min) and post serum treatment ($t=90$ min). E. coli MG1655 was used as a control as it is completely killed by serum. Serum and PBS only samples served as sterility controls. Competitive serum resistance assays were performed in the same manner, except that a 50:50 mixture of wild-type (EC958Alac) and mutant strains were used. Viable counts were performed on MacConkey agar, which allowed the differentiation of EC958Alac (non-lactose fermenter) and the mutant strains.

SDS and NaCl sensitivity assays (MIC)

The MICs of SDS and NaCl were determined by broth microdilution method as previously described [120]. We used five concentrations for SDS including 0.125%, 0.0625%, 0.031%, 0.016% and 0.008% in LB. For NaCl, the range of concentration was 0.8 M, 0.6 M, 0.5 M, 0.4 M and 0.3 M.

LPS gel assay

LPS was extracted from bacterial strains and LPS patterns were determined by Tricine-SDS Polyacrylamide gel electrophoresis (TSDS-PAGE) and visualized by silver staining as previously described [121,122].

Supporting Information

Figure S1 Diagram demonstrating the use of the custom primer with four functional regions for PCR enrichment of transposon derived fragments. (PDF)

Figure S2 Results of edgeR analysis showing logFC of all genes and highlighting genes that satisfy the stringent cut-off for serum resistome. (PDF)

Figure S3 LPS patterns of all defined mutants generated in this study. The LPS pattern of wild-type strain EC958 (WT) was used as reference to identify alteration in LPS of other mutants. (PDF)

Table S1 Summary of transposon insertions in each CDS within the EC958 chromosome. (XLSX)

Table S2 Sequence comparison of EC958 essential genes and genes from 50 E. coli complete genomes. (XLSX)

Table S3 Essential genes unique to EC958. (XLSX)

Table S4 Results of edgeR analysis comparing read counts between control and test samples to identify the serum resistome. (XLSX)

Table S5 Primers used in this study. (XLSX)

Acknowledgments

We would like to thank Gemma Langridge, Keith Turner, Sabine Eckert, Daniel Turner, Philip Hugenholtz, Nicholas West and Brian Forde for expert advice and Leopold Parts for statistical analysis in R.

Author Contributions

Conceived and designed the experiments: MDP SS MT SAB MAS. Performed the experiments: MDP KMP SS SWL LPA MESA VMM. Analyzed the data: MDP DGM SAB MAS. Contributed reagents/materials/analysis tools: VMM MU SAB. Wrote the paper: MDP SS DGM MT MU SAB MAS.


87. Kircher M, Heyn P, Kebö J (2011) Addressing challenges in the production and
analysis of Illumina sequencing data. BMC Genomics 12: 392.
87. Kircher M, Heyn P, Kebö J (2011) Addressing challenges in the production and
analysis of Illumina sequencing data. BMC Genomics 12: 392.
enzymes to change the stereochemical configuration of a single carbon in
alpha-D-glucosyl-1-phosphate uridylyltransferase of Escherichia coli: isolation and
of a knockout library and defined deletion strains of Escherichia coli K-12 on a minimal
92. Barquist L, Boinett C, Gain AK (2013) Approaches to querying bacterial
Experimental determination and system level analysis of essential genes in
of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio
lipopolysaccharide of an extraintestinal isolate of Escherichia coli in serum resistance.
101. Li H, Ruan J, Durbin R (2008) Mapping short DNA sequencing reads and
Accumulation of a polyisoprene-linked amino sugar in polymyxin-resistant
A graphic browser for mapping, visualizing and annotating DNA sequence.
Bioinformatics 16: 944–945.


