Genome-Wide Identification by Transposon Insertion Sequencing of *Escherichia coli* K1 Genes Essential for *in vitro* Growth, Gastrointestinal Colonizing Capacity and Survival in Serum

Alex J. McCarthy¹*, Richard A. Stabler², Peter W. Taylor¹

1 School of Pharmacy, University College London, London, WC1N 1AX, UK
2 London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK

*Correspondence to Peter W. Taylor, peter.taylor@ucl.ac.uk

Short title: Essential *E. coli* K1 genes

*Current address: Bacterial Infections and Immunity, Universitair Medisch Centrum Utrecht, 3584 CX Utrecht, The Netherlands

Word count: Abstract 235; text 8008
Escherichia coli K1 strains are major causative agents of invasive disease of the new born. The age dependency of infection can be reproduced in the neonatal rat. Colonization of the small intestine following oral administration of K1 bacteria leads rapidly to invasion of the blood circulation; bacteria that avoid capture by the mesenteric lymphatic system and evade antibacterial mechanisms in the blood may disseminate to cause organ-specific infections such as meningitis. Some E. coli K1 surface constituents, in particular the polysialic acid capsule, are known to contribute to invasive potential but a comprehensive picture of the factors that determine the fully virulent phenotype has not so far emerged. We constructed a library and constituent sub-libraries of ~775,000 Tn5 transposon mutants of E. coli K1 strain A192PP and employed transposon-directed insertion site sequencing (TraDIS) to identify genes required for fitness for infection in the two-day-old rat. Transposon insertions were lacking in 357 genes following recovery on selective agar; these genes were considered essential for growth in nutrient replete medium. Colonization of the mid-section of the small intestine was facilitated by 167 E. coli K1 gene products. Restricted bacterial translocation across epithelial barriers precluded TraDIS analysis of gut-to-blood and blood-to-brain transits; 97 genes were required for survival in human serum. The study revealed that a large number of bacterial genes, many not previously associated with systemic E. coli K1 infection, are required to realise full invasive potential. Importance Escherichia coli K1 strains cause life-threatening infections in newborn infants. They are acquired from the mother at birth and colonize the small intestine, from where they invade the blood and central nervous system. It is difficult to obtain information from acutely ill patients that shed light on physiological and bacterial factors determining invasive disease. Key aspects of naturally occurring age-dependent human infection can be
reproduced in neonatal rats. Here, we employ transposon-directed insertion site sequencing to identify genes essential for *in vitro* growth of *E. coli* K1 and genes that contribute to colonization of susceptible rats. The presence of bottlenecks to invasion of the blood and cerebrospinal compartments precluded insertion sequencing analysis but we identified genes for survival in serum.
Early-onset sepsis and associated septicemia and meningitis are major causes of morbidity and mortality in the first weeks of life. In the developed world, encapsulated *Escherichia coli* and Group B streptococci are responsible for the large majority of these infections (1-3). Over 80% of *E. coli* blood and cerebrospinal fluid isolates from infected neonates express the α,2-8-linked polysialic acid (polySia) capsular K1 polysaccharide (4, 5), a polymer facilitating evasion of neonatal immune defenses due to its structural similarity to the polySia modulator of neuronal plasticity in the developing human embryo (6). Infections arise due to colonization of the neonatal gastrointestinal (GI) tract by maternally derived *E. coli* K1 at or soon after birth, from where the bacteria invade the systemic circulation to gain entry into the central nervous system (CNS) (7, 8).

Essential features of the human infection can be reproduced in the neonatal rat, enabling investigation of the pathogenesis of *E. coli* K1 neonatal invasive infections (9-11). In susceptible two-day-old (P2) rat pups the protective mucus layer in the small intestine is poorly developed but matures to full thickness over the P2-P9 period, coincident with the development of resistance to invasive infection from GI-colonizing *E. coli* K1 (12). Thus, oral administration of *E. coli* K1 initiates stable colonization of the small intestine in both P2 and P9 pups but elicits lethal systemic infection only in younger animals (13). In the absence of an effective mucus barrier at P2, the colonizing bacteria make contact with the apical surface of enterocytes in the mid-region of the small intestine before translocation to the submucosa by an incompletely defined transcellular pathway (12). They subsequently gain access to the blood compartment by evading mesenteric lymphatic capture (10, 14). *E. coli* K1 cells strongly express polySia in blood (15) and the capsule may protect the bacteria from complement attack during this phase of the infection by facilitating binding of complement.
regulatory factor H to surface-bound C3b to prevent activation of the alternative pathway (16, 17). Following hematogenous spread, the bacteria enter the CNS via the blood-cerebrospinal barrier at the choroid plexus epithelium to colonize the meninges (15). Some microorganisms that invade the CNS enter across the cerebral microvascular endothelium of the arachnoid membrane (18), although the restricted distribution of E. coli K1 within the CNS suggests this is not a primary route of entry for this pathogen.

Only a limited number of pathogenic bacteria have the capacity to invade the CNS from a remote colonizing site and the large majority elaborate a protective capsule that facilitates avoidance of host defenses during transit to the site of infection (19). Although the polySia capsule is clearly necessary for neonatal pathogenesis of E. coli K1 (11), the large majority of bacterial virulence factors that facilitate transit from GI tract to brain are unknown. A number of potential virulence factors associated with neonatal bacterial meningitis have been defined by phylogenetic analysis (20) and there is good evidence that the genotoxin colibactin and the siderophore yersiniabactin contribute to the pathogenesis of E. coli K1 in the experimental rat (21-23); however, a more detailed understanding of virulence mechanisms of E. coli K1 invasive disease will present opportunities for new modes of therapy for these devastating infections.

Transposon insertion sequencing (24, 25), a combination of traditional transposon mutagenesis and massively parallel DNA sequencing, is a powerful tool for the genome-wide enhanced genetic screening of large pools of mutants in a single experiment. It has recently been used to determine the full complement of genes required for expression of the K1 capsule by an E. coli uropathogenic isolate (26). The technique can be used to detect variations in genetic fitness of individual mutants undergoing selection in colonized and infected hosts. There are a number of variations of this procedure but all rely on the
creation of a pool of insertion mutants in which every locus has been disrupted at multiple sites; determination of the site of transposon insertion by sequencing of transposon junctions within chromosomal DNA before and after applying selective pressure will identify mutants attenuated under the selective condition (27). Thus, genes that confer fitness during *Klebsiella pneumoniae* (28) and *Acinetobacter baumannii* (29) lung persistence, systemic and mucosal survival of *Pseudomonas aeruginosa* (30), and spleen colonization in the mouse of uropathogenic *E. coli* (31) have been identified by this approach. In this study, we employ transposon-directed insertion site sequencing (TraDIS) (24) to interrogate a library of ~775,000 Tn5 mutants or constituent sub-libraries of *E. coli* K1 strain A192PP for genes essential for growth in vitro and GI colonization, invasion and systemic survival in susceptible P2 rat pups. In addition, we identified “bottlenecks” (32) to systemic invasion that restrict population diversity and limit the potential for transposon insertion site analysis of infection in the GI-colonized neonatal rat.

**RESULTS**

**Generation of a Tn5 mutant library and identification of essential genes.** To provide sufficient saturation density for the identification of *E. coli* K1 genes essential for growth in vitro and of those conferring fitness in a range of defined environments, approximately 300 individual pools, each of 1-5 x 10^3 transposon mutants of *E. coli* A192PP, were constructed and combined to form a library containing over 7.75 x 10^5 mutants. Linker PCR was performed on randomly selected mutants to confirm that Tn5 had inserted into random genomic locations (Fig. S1). TraDIS was performed on pooled but uncultured mutants to identify Tn5 insertion sites within the 5.52 Mbp genome of A192PP (33). Sequences of indexed amplicons were determined and 2 x 10^6 sequence reads containing Tn5 were
mapped onto the *E. coli* K1 A192PP genome. Reads mapped to 237,860 unique Tn5 insertion sites, and were distributed along the entire genome (Fig. 1A).

As the Tn5 library contained a high transposon insertion density, genes with no or limited Tn5 insertion sites are likely to be essential for growth in nutrient-replete media such as Luria-Bertani (LB) broth. We calculated insertion indices for each gene by normalizing the number of insertions in each gene by gene length. Insertion index values for two technical replicates were highly correlated (Spearman’s rho = 0.9589) (Fig. 1B). A density plot of insertion indices produced a bimodal distribution, with a narrow peak representing genes with no or a limited number of Tn5 insertions and a broad peak containing genes with a large number of Tn5 insertions (Fig. 1C); the former comprised genes that confer lethality when mutated and the latter genes that can be mutated without affecting bacterial viability. To identify genes significantly lacking Tn5 insertions and therefore essential for *in vitro* growth, gamma distributions from the density plot were used to determine log₂ likelihood ratios. Examples of essential genes containing no or limited Tn5 insertions are shown in Fig. 1D. A total of 357 genes were predicted to be essential for the *in vitro* growth of *E. coli* K1 A192PP and these are shown in Table S1, together with KEGG (Kyoto Encyclopedia of Genes and Genomes) descriptors for genes involved in metabolic pathways.

COG (Clusters of Orthologous Groups) was used to identify the functional category of each gene essential for growth *in vitro* from the A192PP whole genome sequence (accession number PRJEB9141). Genes involved in ribosomal structure (11% of total number of essential genes) and protein biosynthesis (15%) featured prominently and were significantly enriched in relation to their representation within the whole genome as were genes encoding proteins for DNA replication (3%), cell wall (peptidoglycan, lipopolysaccharide)
biosynthesis (6.25%) and membrane biogenesis (3%) (Fig. 2). Genes for protein secretion and export as well as ABC transporter genes were also well represented; the remaining essential genes were involved in a wide variety of cellular catabolic and anabolic functions. The list features 254 genes that were found by TraDIS (34) to be essential for growth in Luria broth of an *E. coli* ST131 multi-drug resistant urinary tract isolate (from a total of 315 essential genes). In similar fashion, 253 genes determined as essential for growth of *E. coli* K12 MG1655 in LB broth were also identified as essential in the current study (Table S1); the K12 study employed a comprehensive set of precisely defined, in-frame single-gene deletion mutants (35), not transposon insertion sequencing. 

**Maintaining Tn5 library diversity.** The polySia capsule is a major determinant of virulence in *E. coli* K1 and is central to the capacity of K1 clones to cause neonatal systemic infection (11, 36). PolySia biosynthesis imposes a substantial metabolic burden on producer strains (37). As TraDIS and other transposon insertion sequencing procedures generally employ growth in liquid medium for recovery and expansion of the output pool (38), we investigated the impact of batch culture on the expression of the K1 capsule within the Tn5 library. The complete Tn5 library was inoculated into LB broth, incubated for 8 h at 37°C and the proportion of encapsulated and non-encapsulated A192PP bacteria determined by susceptibility to the *E. coli* K1-specific bacteriophage K1E within the population. Non-encapsulated mutants initially comprised 4.66% of the bacterial population but by the end of the incubation period this had risen to 98.24% (Fig. 3A). Growth rates in LB broth of *E. coli* A192PP and a non-encapsulated mutant of A192PP randomly selected from the Tn5 library did not differ significantly (Fig. S2A).
The cultured Tn5 library was avirulent as determined by administration to P2 neonatal rat pups whereas GI colonization with 2-6 x 10^6 CFU *E. coli* A192PP and the uncultured Tn5 library were lethal. A similar colonizing inoculum of the cultured (8 h; 37°C) *E. coli* A192PP-Tn5 library had no impact on survival and all pups remained healthy over the seven-day observation period (Fig. 3B), even though all animals remained heavily colonized with K1 bacteria throughout the experiment (data not shown). Thus, culture of the library prior to challenge resulted in loss of phenotypic diversity and virulence. The complete Tn5 library contained 2.81 x 10^5 unique Tn5 insertions, of which 750 (2.66% of the bacterial population) possessed transposon insertions in genes determining capsule biosynthesis (data not shown). The probability that cultured sub-libraries of more than 5 x 10^3 mutants contained a non-capsulated mutant was calculated to be ≥0.98 but only 0.55 for sub-libraries of 1 x 10^3. Low complexity libraries of 10^3 mutants maintained virulence in P2 neonatal rat pups after culture whereas more complex libraries did not (Fig. S3), due to the absence of mutants lacking the capacity to express the polySia capsule within the inoculum. To minimise bias, in all subsequent experiments libraries of sufficient complexity to contain multiple numbers of non-encapsulated mutants were used; for experiments utilizing neonatal rats the period between colonization initiation and tissue harvesting was kept to a minimum and tissue homogenates were cultured directly on to selective agar plates with no intervening liquid culture step.

**Genes required for GI colonization.** *E. coli* A192PP colonize the small intestine of neonatal rats following oral administration of the bacterial bolus, with 10^7-10^8 K1 bacteria/g intestinal tissue persisting for at least one week (12, 13). Translocation of the neonatal pathogen to the blood compartment *via* the mesenteric lymphatic system occurs predominantly, and in
all likelihood exclusively, across the epithelium of the mid-section of the small intestine (MSI), even though the density of colonizing bacteria in this region of the GI tract is no greater than that within neighbouring proximal (PSI) or distal (DSI) locations (12).

Few attempts have been made to determine the genes or gene products required by *E. coli* K1 for colonization of the GI tract (39). To prevent loss of diversity of the *E. coli* K1 A192PP-Tn5 library, we minimized the period of colonization before sampling the *E. coli* K1 population of the MSI. The colonizing *E. coli* K1 population in proximal, middle and distal regions of the small intestine did not expand beyond 4 h after initiation of colonization (Fig. 4A); GI tissues were therefore excised at this time point. To identify mutants with decreased capacity to colonize the MSI, P2 rats were fed 1 x 10⁹ CFU of an *E. coli* K1 A192PP-Tn5 library containing 2 x 10⁵ mutants, the pups sacrificed after 4 h and *E. coli* K1 bacteria in the MSI enumerated. The bacterial load of rats colonized with the Tn5 library was comparable to that of rats colonized with the wildtype strain (data not shown). MSI tissues from four rats were pooled, homogenized and cultured on LB agar containing kanamycin to ensure that mutant frequency was not overestimated by inclusion of measurements of DNA from dead bacteria; kan⁺ colonies were then pooled, DNA extracted and the fitness of each mutant determined by TraDIS. Input and output pools each comprised 2 x 10⁵ CFU and the ratio of input:MSI read counts were expressed as log₂ fold change. A wide distribution of fitness scores (40) were detected (Fig. 4B). The majority of transposon insertions did not have a strong negative or positive effect on colonisation of the MSI. A total of 387 transposon insertions, within 167 genes, had significantly decreased in normalized read counts between input and output pools (negative log₂ fold change and \( P < 0.05 \); Table S2). Of the 387 insertion sites, 180 were not detectable in the output pool, demonstrating complete loss in the output pool. Many of these transposon insertion sites occurred within the same gene.
For example, within the *neuC* gene, 70 unique transposon insertion sites were identified as lost during colonization. Transposon-interrupted genes were identified as important for colonization of the MSI and were grouped into seven arbitrary categories: (i) genes encoding surface structures, including pili, (ii) genes encoding secretory components, (iii) genes involved in intermediary metabolism, (iv) stress response genes, (v) cytoplasmic membrane (CM)-located genes, (vi) genes for iron acquisition and (vii) others and hypothetical genes.

A high proportion of mutations associated with decreased MSI colonizing capacity were located in genes affecting the biosynthesis of surface structures (Table S2). A few genes were involved in lipopolysaccharide (LPS) biosynthesis (*yrbH, yiaH*) and OM proteins (*ompG, ycbS*) but the majority affected the polySia capsule, with genes of the *neu* operon (41), accounting for 194 of the 387 colonization-attenuated mutants. There is some evidence that capsular polysaccharides may promote adhesion to biological and non-biological surfaces during biofilm formation (37) but there has been little or no consideration of a role for capsules as mediators of GI colonization.

A limited number of genes associated with type II and IV secretion were identified as required for colonisation of the MSI; these multiprotein complexes translocate a wide range of proteins and protein complexes across host membranes (42, 43) and are implicated in adherence and intestinal colonization of enterohemorrhagic *E. coli* in farm animals (44).

Genes for assembly of pilus proteins, including some encoded by the *tra* locus, likely to be located on plasmids, that initiate conjugation, were also linked to colonization; pili are virulence factors that may mediate attachment to and infection of host cells (45).

Colonization by both commensals and pathogens is dependent on nutrient scavenging, sensing chemical signals and regulation of gene expression as the bacteria adapt to a new environment.
and potentially hostile environment that in the case of *E. coli* K1 appears to rely on stress response genes such as *yhiM* (encodes a protein aiding survival at low pH) and the heat shock protein genes *clpB* and *yrfH*, as well as DNA repair genes. A large number encoded enzymes involved in the metabolism of sugars (e.g., *gcd*, *rpiR*, *glgC*), amino acids (*dadX*, *metB*, *tdcB*), fatty acids (*yafH*, *fixA*), growth factors (*bisC*, *yigB*, *thiF*) and other secondary metabolites (*yicP*). Transporters and permeases involved in central intermediary metabolism also featured prominently: these included permeases of the major facilitator superfamily (*YjiZ*), the hexose phosphate transport protein *UhpT*, the carnitine transporter *CaiT* and a range of CM-located sugar transporters. Of note was the impact of mutation of the *fucR* L-fucose operon activator on colonization; fucose is abundant in the GI tract and the fucose-sensing system in enterohemorrhagic *E. coli* regulates colonization and controls expression of virulence and metabolic genes (46). Availability of free iron is severely limited in the GI tract and ingestion of iron predisposes to infection (47); the importance of iron acquisition for *E. coli* K1 during GI colonization is reflected in the requirement for a number of genes related to iron uptake (e.g., *feoB*, *fepA*).

**GI colonizing capacity and virulence of single gene mutants.** To investigate the contribution of the polySia capsule to colonization of the neonatal rat GI tract, we disrupted the *neuC* gene of *E. coli* A192PP genes using bacteriophage λ Red recombinase to produce a capsule free mutant as judged by resistance to *E. coli* K1-specific phage K1E. We also produced other single gene mutants in genes identified by the TraDIS GI screen: *vasL* (encoding a type IV secretion system protein), *yfeC* (predicted to form part of a toxin/anti-toxin locus) and two genes with unknown function, *yaeQ* and *A192PP_3010* (the latter is present in genomes of other extra-intestinal *E. coli* pathogens, including IHE3034, UTI89, RS218, PMV-1 and S88).
Growth rates of these mutants, in particular the capsule-negative neuC mutant (Fig. S2B), were indistinguishable from that of the *E. coli* A192PP parent in LB medium. All were examined for their capacity to colonize the GI tract and cause lethal infection in P2 rat pups (Fig. 4C and 4D).

The *E. coli* A192PP parent strain or single gene mutants (2-6x10⁶) were administered orally to P2 rats; all members of a litter of 12 pups received the same strain. Pups were sacrificed 24 h after initiation of colonization and *E. coli* K1 bacteria in the small intestine (PSI, MSI and DSI) and colon enumerated. The capacity of all mutants to transit the upper portion of the alimentary canal, pass through the stomach and colonize the small intestine was markedly inferior to the wildtype strain (Fig. 4C). Reductions in colonization of the PSI, MSI and DSI by the mutants, including *E. coli* A192PPΔneuC::kan, were significant, the only exception being colonization of the DSI by A192PPΔyfeC::kan, with no significant difference between parent and mutant. Interestingly, no increases in the numbers of viable A192PPΔneuC::kan, A192PPΔvasL::kan, A192PPΔ3010::kan and A192PPΔyaeQ::kan recovered from the colon were noted to compensate for reductions in colonization of the small intestine. There was a significant increase in the colonic burden of viable A192PPΔyfeC::kan bacteria compared to the parent strain. We have established (12) that *E. coli* A192PP transits to the blood circulation via the mesenteric lymphatic system by exploiting a vesicular pathway through the GI epithelium only at the MSI. As mutant numbers colonizing this region of the small intestine were much reduced compared to the parent strain, we determined the capacity of the single gene mutants to elicit lethal systemic infection following GI colonization by oral administration of 2-6x10⁶ bacteria at P2 (Fig. 4D). Four of the five mutants (A192PPΔneuC::kan, A192PPΔvasL::kan, A192PPΔ3010::kan and A192PPΔyfeC::kan) displayed significantly reduced lethal potential.
compared to the A192PP parent. Loss of capsule (A192PPΔneuC::kan) resulted in complete loss of lethality over the seven day observation period. Administration of A192PPΔvasL::kan elicited a lethal response in 41.6% of pups; 33.3% and 25% survived after receiving, respectively, A192PPΔ3010::kan and A192PPΔyfeC::kan at P2. For A192PPΔyaeQ::kan, 75% of pups succumbed to lethal infection but did not reach levels of significance when compared to the 100% lethality engendered by the A192PP parent (P >0.05). Overall, these data indicate that the TraDIS screen efficiently identified genes important for MSI colonization that impact on pathogenic potential.

A bottleneck to infection in the neonatal rat prevents identification of genes for translocation across the gastrointestinal epithelium. Our initial intention was to exploit the high degree of susceptibility of the P2 neonatal rat to systemic infection, sepsis and meningitis following oral administration of an effective dose of E. coli A192PP in order to determine all genes required to enable the neonatal pathogen to overcome previously defined (12-15) physical and immunological barriers to invasion of the blood circulation and dissemination to the meninges. However, earlier studies indicate that relatively few E. coli K1 bacteria migrate from colonized sites within the GI tract to the blood (10), constraining the genetic diversity of the translocated bacterial population and eliminating genotypes from the translocated gene pool in a stochastic manner that does not reflect the fitness of individual genes to contribute to genotypes with invasive potential (32). We therefore determined if bottlenecks existed which would compromise the identification of mutants with attenuated capacity to translocate from the GI tract to the blood compartment; if any experimental bottlenecks are narrower than the complexity of the E. coli A192PP Tn5 library, many relevant transposon insertion mutants will be lost entirely by chance (38).
Further, the existence of a restrictive bottleneck would limit the complexity of the library that could be used for TraDIS evaluation of populations colonizing the MSI (input pool) and reaching the blood (output pool).

We constructed an *E. coli* A192PPΔlacZ mutant by bacteriophage λ Red recombineering and confirmed that there was no significant difference in lethal potential between *E. coli* A192PP and the lacZ mutant (Fig. 5A). We then used mixtures of parent and mutant to investigate the existence of bottlenecks that restrict translocation to the blood compartment. A 1:1 mixture (total 2-4 x10⁶ CFU) of *E. coli* A192PP and A192PPΔlacZ was administered orally to P2 rat pups, the animals sacrificed after 24 h and GI tissue homogenates plated for quantification of each strain. The competitive index (CI), the ratio of input A192PP: A192PPΔlacZ to output A192PP: A192PPΔlacZ, was calculated for excised PSI, MSI, DSI, colon and mesenteric lymphatic tissue and for blood. CI values in the PSI, MSI, DSI and colon were not significantly different from 1 (one-sample t-test), indicating that the composition of the colonizing inoculum was maintained in each rat pup (Fig. 5B). However, there was more heterogeneity in CI values of bacterial populations from the blood and in five pups only one strain could be recovered from the blood (four animals parent strain only, one animal A192PPΔlacZ only). The highly restrictive bottleneck between GI epithelial transport and entry into the blood circulation supports the argument that reduced virulence of the complete, cultured library in comparison to less complex sub-libraries (Fig. S3) is at least in part due to a reduced likelihood that a fully virulent mutant would randomly escape capture by the mesenteric lymphatic system. The presence of significant bottlenecks between the GI tract, blood circulation and brain was confirmed by determination of the complexity of recovered Tn-5 library populations from these sources (Fig. 5C).
Identification of *E. coli* K1 A192PP genes required for survival in human serum. Systemic infection in the neonatal rat is likely to be maintained only if *E. coli* A192PP bacteria survive in the blood circulation. Due to limited exposure to antigens *in utero* coupled with deficits in adaptive immunity, neonates depend on innate immunity for protection against infection. The complement system provides front line innate defense against Gram-negative bacterial infection and the polySia capsule in turn enables *E. coli* K1 to avoid successful complement-mediated attack by host immune mechanisms. To obtain insights into *E. coli* K1 pathogenesis during the invasive phase of the infection, and in light of restrictions placed on the neonatal rat model with regard to the use of TraDIS by the gut-to-blood bottleneck, we used the *E. coli* A192PP Tn5 library to investigate genes essential for A192PP fitness in pooled normal human serum, a reliable and plentiful source of all soluble components of the three complement pathways (48).

*E. coli* A192PP is resistant to the bactericidal action of human serum (Fig. 6A). A portion of the A192PP-Tn5 library containing $2 \times 10^9$ mutants ($1 \times 10^9$ CFU) was incubated in either 30% human serum or 30% heat-inactivated serum (final volume 375 µl) at 37°C for 3 h. Kan$^R$ bacteria in the input and output pools (each $2 \times 10^5$) were collected, DNA extracted from each pool and transposon insertion sites sequenced. A wide distribution of fitness scores were detected (Fig. 6B). Mutation of 97 genes (negative log$_2$-fold change and $P<0.05$) resulted in decreased survival in normal serum, but not in heat-inactivated serum (Fig. 6C & Table S3).

A high proportion of genes identified in the TraDIS screen as contributing to resistance encoded cell surface constituents. It is well established that the polySia capsule protects *E. coli* K1 from complement attack (16, 17) and three mutations in the *kps* capsule gene cluster compromised serum survival. The central region of the cluster contains the *neu*
genes that direct the biosynthesis, activation and polymerization of the N-acetylneuraminic acid building block of polySia. neuC encodes the UDP N-acetylglucosamine 2-epimerase that catalyzes the formation of N-acetylmannosamine (49) and the O-acetyltransferase neuD acetylates monomeric neuraminic acid at carbon position 7 or 9 (50). KpsM is a component of the multimeric ATP-binding cassette transporter involved in the translocation of the polySia capsule through a transmembrane corridor to the cell surface (41, 51). Disruption of the genes encoding these proteins will prevent polySia expression (41); interruption of rfaH, identified in the TraDIS screen, will also prevent capsule expression but its loss will have a more profound effect on the surface topography of E. coli A192PP, as this transcriptional anti-terminator is required for the expression of operons that direct the synthesis, assembly and export of LPS core components, pili and toxins in addition to the capsule (52, 53).

Indeed, survival in serum is dependent on anti-termination control by RfaH (54). Another gene identified that impacts on capsule formation was bipA; BipA is a tyrosine-phosphorylated GTPase that regulates through the ribosome a variety of cell processes, including some associated with virulence (55, 56). Other genes involved in LPS biosynthesis and pilus formation were also identified: waaW is a UDP-galactose:(galactosyl) LPS alpha1,2-galactosyltransferase involved in the synthesis of the R1 and R4 LPS core oligosaccharides (57) and wzzE encodes a polysaccharide copolymerase that catalyzes the polymerization of LPS O-antigen oligosaccharide repeat units into a mature polymer within the periplasmic space in readiness for export to the cell surface (58). Both mutations will prevent attachment of LPS O-side chains to the core oligosaccharide of LPS. The 16 genes that specify pilus synthesis that were identified in the screen included the majority of genes of the tra locus.
The TraDIS screen identified a range of proteins that are embedded in the OM (Fig. 391C), none of which had been previously implicated in complement resistance, and which could influence the topography of the bacterial surface. Of the remaining genes with assigned function, the majority were involved with cell metabolism and the stress response; it is well established that metabolic processes are intimately associated with the complement-mediated bacterial killing process (59, 60). To verify the screen, we constructed four single gene mutants of *E. coli* A192PP by bacteriophage λ Red recombineering. Genes with roles in LPS synthesis (*rfaH* and *waaW*), capsule synthesis (*neuC*) and pilus assembly (*traL*) were mutated; none showed any reduction in growth rate in LB broth. All displayed significant reductions in complement resistance following incubation in pooled human serum (Fig. 6D). *E. coli* A192PPΔ*rfaH* was exquisitely susceptible with no colonies detected after 30 min. The viability of A192PPΔ*neuC* was also compromised with a threefold log reduction in viability over the 3 h incubation period. Killing of A192PPΔ*traL* and A192PPΔ*waaW* was less marked but these mutations significantly reduced viability. Complementation of the mutants with the functional gene introduced on a pUC19 vector completely restored resistance in all cases (Fig. 6D). These genes also contributed to lethality in the P2 neonatal rat (Fig. 6E). The lethal capacity of A192PPΔ*neuC*, A192PPΔ*rfaH* and A192PPΔ*waaW* was completely attenuated in comparison to *E. coli* A192PP; 42% of pups administered A192PPΔ*traL* succumbed to systemic infection (all *P* < 0.01).

**DISCUSSION**

Systemic infection with meningeal involvement arises spontaneously after GI colonization of neonatal rats with a high proportion of *E. coli* K1 isolates and the pathway to infection...
mirrors to a large extent that of natural infections in the human host. In contrast to models of bacterial infection that create an artificial pathogenesis bypassing some or all of the barriers to infection by injection of a bacterial bolus directly into the blood circulation, the neonatal rat model provides an opportunity to investigate in stepwise fashion the progress of the pathogen as it transits from gut to blood to brain. TraDIS and other transposon sequencing methods enable simultaneous and rapid determination of the fitness contribution of every gene for a given condition and therefore have the potential to enable the identification of genes that are essential for, or significantly contribute to, each step of the infection process. However, stochastic loss will become evident if each mutant in the input pool does not have an equal chance to overcome the physical, physiological and immunological barriers presented by the host (61). This was clearly the case with epithelial transit of *E. coli* A192PP, with evidence that on occasion systemic infection arose due to only one viable bacterial cell entering the blood circulation (Fig. 5B), and complements other studies showing single or low-cell-number bottlenecks in models of severe infection (62-64). As translocation from MSI colonizing sites to the blood was not amenable to analysis by TraDIS we determined genes essential for survival in the presence of complement, a major component of the innate immune system that protects against extracellular systemic pathogens (17).

The high density of transposon insertion into random genomic positions along the entire *E. coli* A192PP chromosome, with minimal insertional bias (Fig. 1A), enabled the identification of genes essential for growth in nutrient replete LB medium. Of the 357 *E. coli* A192PP genes considered essential, orthologues of 254 (from 315) had been previously identified using TraDIS in a multi-drug-resistant uropathogenic strain of *E. coli* ST131 grown in LB (34) and 253 in an *E. coli* K12 strain (35), confirming the existence of a core set of
essential genes in *E. coli*. As anticipated, a high proportion of these genes encoded enzymes involved in a range of key metabolic functions such as carbohydrate, protein and nucleobase metabolism, and the remainder were associated with essential functions such as transport, cell organisation and biogenesis.

During characterization of the *E. coli* A192PP mutant library we examined the impact of culture in liquid medium on the expression of the polySia capsule, which places large demands on cell energy expenditure, as lengthy incubation times before marker selection may lower library complexity (38). Unexpectedly, we found that prolonged culture of the library enriched the proportion of non-encapsulated mutants (Fig. 3A). We anticipated that loss of capsule would enable the non-encapsulated mutants to grow at a faster rate than capsule-replete mutants and wildtype, and out-compete capsule-bearing library members. However, growth of a non-encapsulated mutant selected at random from the library was virtually identical to, and not significantly different from, the *E. coli* A192PP parent strain (Fig. S2A). There was also no difference in the climax populations of the strains at the end of the logarithmic phase of growth. In similar fashion, the growth curve for a *neuC* single gene mutant was identical to *E. coli* A192PP (Fig. S2B). *neuC* is involved in the synthesis of the N-acetylneuraminic acid monomeric unit of polySia, and as a consequence is unable to elaborate the capsule. It is clearly impractical to evaluate the growth kinetics of every distinct non-encapsulated mutant in the Tn5 library but it currently appears that differences in growth rate of individual library members cannot explain the highly reproducible enrichment that we observed. Indeed, use of transposon insertion libraries is predicated on the assumption that there are no significant differences in the growth rate of individual mutants. At present, the basis of the loss of mutants expressing capsule in TraDIS library cultures cannot be readily explained.
A sub-library of $2 \times 10^5$ mutants was used to establish genes involved in GI colonization. To minimize bias due to any outgrowth of non-encapsulated mutants on the GI epithelium we harvested *E. coli* K1 from the MSI after 4 h, by which time maximal CFU had been achieved; bacteria were plated directly on to solid medium to further avoid outgrowth. Bias due to this restricted timeline is likely to be low as the majority of genes involved in adhesion and complement resistance are expressed constitutively. TraDIS identified the polySia capsule as a major determinant of GI colonization associated with *E. coli* K1. There is little or no evidence from the literature that capsules of Gram-negative bacteria enhance GI colonization; indeed, it has been reported that they interfere with adhesive interactions by obstructing binding of underlying surface molecules to mucosal surfaces (65, 66). The single gene mutant *E. coli* A192PPΔneuC::kan displayed a reduced capacity to colonize the MSI (Fig. 4E), although it should be borne in mind that passage through the upper alimentary canal and stomach may impact on the number of mutant bacteria gaining access to the small intestine. In this context it should be noted that capsular exopolysaccharide protects *E. coli* from the environmental stress of stomach acid (67).

Other cell surface structures that are likely to have an impact on adhesion and colonization of the mucosal layer associated with the MSI were identified by TraDIS. Pili are established mediators of adhesion of *E. coli* to the host epithelium, although a large proportion of the evidence comes from enterotoxigenic and enteropathogenic strains (68, 69). LPS and OM protein encoding genes were also implicated, as were genes involved in the stress response, reflecting ongoing adaptation to a new and hostile environment. The involvement of genes encoding metabolic enzymes, including some for anaerobic respiration, equates to increases in bacterial cell numbers in the anaerobic environment of the small intestine and for iron acquisition genes this reflects the low availability of...
intestinal luminal iron (47, 70). Genes encoding some components of type II and type IV secretion systems were found with decreased frequency in the output pool. Members of these gene categories were also identified by Martindale et al (39) as necessary for GI colonization of *E. coli* K1 faecal isolate RS228 using signature-tagged mutagenesis; no genes found in this study were identified in the current study, in spite of the close genetic relatedness of the strains employed.

The intestinal lumen represents a potentially important portal of entry for pathogens into the host through adhesion, invasion or disruption of the epithelial barrier (71). In newborns, *E. coli* K1 induces no detectable disruption of barrier integrity but exploits an intracellular pathway to access the submucosa (12). Only small numbers of bacteria breach the mesenteric lymphatic barrier in apparently random fashion (Fig. 5) and this precludes analysis by TraDIS. To accumulate data on genes and gene products facilitating invasion and survival/replication in the blood circulation, we examined essentiality for avoiding complement-mediated bactericidal effects. Although not all *E. coli* K1 isolates from cases of systemic infection are resistant to complement, resistance amongst K1 and K5 capsular types is more frequently encountered that for other K types (72); *E. coli* O18:K1 strains (such as A192) are in turn more often resistant than other O:K serotype combinations (73) due to the capacity of the polySia capsule to prevent complement activation. It is assumed, but not established, that the polySia capsule surrounding susceptible strains does not completely mask either OM-located activators of complement or lipid domains on the outer surface of the cell that are targets for OM intercalation of the C5b-9 membrane attack complex, the entity responsible for bacterial killing (59). In addition, long and numerous LPS O-side chains are necessary but not sufficient to enable the target cell to avoid complement killing (74) and they are able to bind C1 inhibitor to arrest classical or lectin pathway activation at the
early C1 stage (75). The importance of these structures for the complement resistance of E. coli K1 is supported by the decreased frequency of key LPS and capsule genes in the output pool along with a large number of OM-embedded proteins.

A small number of OM proteins, such as TraT and Iss, have been implicated as determinants of complement resistance (74) but they have been introduced into low-resistance backgrounds in high copy number; their role in the intrinsic resistance of clinical isolates is unclear and no mechanisms have been invoked to account for increases in resistance. The insertion of large numbers of protein molecules into the OM may fortuitously alter the biophysical properties of the bilayer, reducing the surface area and fluidity of lipid patches that are essential for binding and assembly of the C5b-9 membrane attack complex. The identification by TraDIS of a range of OM proteins as putative complement resistance determinants creates an opportunity to systematically investigate their precise function through generation of single gene mutants and we intend to pursue this line of investigation. We suggest that the architecture of the external surface of the OM, together with other more external macromolecular structures such as polysaccharide capsules, influences the capacity of the pore-generating C5b-9 complex to perturb the integrity of the OM. Thus, the surface of susceptible strains contains a sufficient number of exposed lipid domains to facilitate C5b-9 generation and penetration whereas the spatial and temporal organization of the OM of resistant bacteria is dominated by supramolecular protein assemblages to a degree where insufficient hydrophobic domains are available to act as C5b-9 assembly and binding sites, and this state persists throughout the growth cycle. The data we have generated in this study is compatible with this hypothesis. An array of metabolic genes emerged as essential for maintenance of the complement resistant phenotype (Fig. 6D) and may be indicative of repair processes invoked due to complement
attack. Exposure of resistant *E. coli* to complement results in minor perturbation of membrane integrity and metabolic homeostasis (76, 77) and C5b-9 intercalation into the OM has profound effects on cellular metabolic parameters (60).

TraDIS has also been employed by Schembri and coworkers to define the serum resistome of a globally disseminated, multidrug resistant clone of *E. coli* ST131 (34). They identified, and in most cases validated, 56 genes that contributed to the high level of complement resistance displayed by this pathogen. In similar fashion to our study, genes involved in the synthesis and expression of cell surface components were prominent. A number of genes contributing to LPS biosynthesis such as those of the *waa* operon, the *wzz* locus and *rfaH*, were common to both studies, as was the gene encoding the intermembrane protein AcrA. Genes of the plasmid-encoded *tra* locus, which we determined to be components of the *E. coli* A192PP serum resistome, were not present in *E. coli* ST131 (34) but other OM-located proteins may fulfill a similar role in reducing the fluidic properties of the bilayer. In contrast to the well-established role of the *E. coli* K1 polysialyl polymer in prevention of complement activation, no capsule genes were identified as components of the serum resistome of *E. coli* ST131, but different ST131 isolates express different capsule types due to extensive mosaicism at the capsule locus (78) and these uronic acid-containing polymers are unlikely to prevent complement activation (75). Thus, the different strategies employed by the two strains to prevent successful complement attack, together with differences in the bacterial surface composition and topography, probably explain variations in the serum resistomes of these related pathogens.

In summary, we identified *E. coli* K1 genes required for growth in standard laboratory liquid medium and for colonization of the GI tract of P2 neonatal rat pups. Both data sets provide insights into the biology of K1 neuropathogens and could provide the basis
for drug discovery programs for identification of selective antibacterial or colonization-inhibiting agents. In our rodent model, the stochastic nature of invasion of blood and probably brain prevented TraDIS analysis of gene essentiality for crossing gut epithelial and choroid plexus borders but some indication of genes necessary for survival in blood were obtained from output pool analyses after incubation of *E. coli* A192PP in human serum, a potent source of complement.

**MATERIALS AND METHODS**

**Ethics statement.** Animal experiments were approved by the Ethical Committee of the UCL School of Pharmacy and the United Kingdom Home Office and were conducted in accordance with national legislation.

**Bacteria and culture conditions.** *E. coli* strain A192PP was obtained by serial passage in P2 neonatal rats of *E. coli* A192 (serotype O18:K1) isolated from a patient with septicemia (79), as described earlier (11). Carriage of the polysialyl K1 capsule was determined with phage K1E (80): colonies were streaked onto MH agar, 10 µl of ~10⁹ PFU/ml phage suspension dropped on each streak, the plates incubated overnight at 37°C and the proportion of encapsulated bacteria within cultures quantified by comparing the ratio of phage-susceptible and phage-resistant colonies. *E. coli* A192PP single gene mutants (Table 1) were constructed using bacteriophage λ Red recombination (81); the oligonucleotides employed for construction of targeted mutants, for confirmation of targeted mutants and for construction of complemented mutants are shown in Tables S3-S5. All were cultured in Luria-Bertani (LB) and on LB agar at 37°C; media were supplemented with either 100 µg/ml ampicillin or 50 µg/ml kanamycin as required.
Tn5 library construction. The EZ-Tn5 <KAN-2> Tnp transposome (Epicentre Biotechnologies) was introduced into E. coli A192PP by electroporation. Transformants were selected by overnight growth on LB plates containing 50 µg/ml kanamycin. Pools of 1-5 x 10^3 colonies were collected and frozen at -80°C in PBS containing 20% glycerol. Aliquots of individual pools were combined to create larger populations of mutants of up to 7.75 x 10^5. Genomic DNA was extracted from 1 ml cultures using the PurElute Bacterial Genomic Kit (Edge Biosystems) following standard protocol.

Linker PCR of Tn5 insertion sites. Linker PCR was used to test individual transformant colonies and to confirm individual random-insertion events. DNA (2.5 µg) was digested with AluI restriction enzyme (Promega) and purified using MinElute PCR purification kit (QIAGen). A linker, formed by annealing of oligonucleotides 254 (5’CGACTGGACCTGGA3’) and 256 (5’GATAAGCAGGGATCGGAACCTCCAGGTCCAGTCG3’), was ligated to purified fragments (50 ng) with Quick Ligation kit (NEB). Linker PCR was performed with linker- and transposon-specific oligonucleotides (258 5’GATAAGCAGGGATCGGAACC3’ and 5’GCAATGTAACATCAGAGATTTTGAG3’ respectively) using HotStart Taq Mastermix kit (QIAGen) and thermocycling conditions of 95°C for 5 min, 35 cycles of 94°C for 45 s, 56°C for 1 min and 72°C for 1 min, and 72°C for 10 min. Resulting amplicons were separated on 1.5% agarose gels at 100 V for 60 min.

Illumina sequencing. For sequencing of Tn5 insertion sites, approximately 2 µg of genomic DNA was degraded to ~ 500 bp fragments by ultrasonication using a Covaris instrument. Fragments were end-repaired and A-tailed using the NEBNext DNA library preparation reagent kit for Illumina sequencing (NEB). Adapters Ind_Ad_T (ACACTCTTTCCCTACACGACGCTCTTCCGATC*T; where * indicates phosphorothionate) and
Ind_Ad_B (pGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTC) were annealed and ligated to DNA fragments. PCR was performed with transposon- and adapter-specific primers Tn-FO

\[ 5'\text{TCGTCGGCAGCGATGTATAAGAGACACGCTCTTCTAGATGACCTGACCTC} \]

and Tn-RO

\[ 5'\text{GTCTCTGCTTCGAGGTGTATAAGACACAGCTCTTCCGATC} \]

`. Tn-FO and Adapt-RO contain a forward overhang and reverse overhang for indexing of amplicons by Nextera index primers (Illumina). PCR was performed using HotStart Taq Mastermix kit (QiAgen) and thermocycling conditions of 95°C for 5 min, 22 cycles of 94°C for 45 s, 56°C for 1 min and 72°C for 1 min, and 72°C for 10 min. Resulting amplicons were separated on 1.5% agarose gels at 70 V for 90 min, and those between 150 and 700 bp selected and purified using QiAquick Gel Extraction kit (QiAgen). Samples were indexed with oligonucleotides from Nextera XT Index Kit (Illumina) using HotStart ReadyMix (Kapa Biosystems) and thermocycling conditions of 95°C for 3 min, 8 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 30 s, and 72°C for 5 min applied. Indexed amplicons were purified using the AMPure XP system (Agencourt). The final concentration of samples was confirmed using Qubit dsDNA BR assays (Thermofisher Scientific). Indexed amplicons were sequenced on an Illumina Mi-Seq platform as 151-bp paired-end reads following manufacturer’s protocol (Illumina).

**Bioinformatic and statistical analysis.** Raw sequence reads that passed Trimmomatic quality control filters (82) and contained the Tn5 transposon were mapped to the *E. coli* K1 A192PP reference genome (14) using Bowtie (83), permitting zero mismatches and excluding reads that did not map to a single site. The reference genome assembly contains ORFs located on...
contigs that were mapped to the IHE3034 chromosome and ORFs located on other contigs that are likely to map to plasmids and other mobile genetic elements. An in-house pipeline based on the SAMtools (http://samtools.sourceforge.net) and BCFtools toolkits was utilised on the alignment files to determine insertion sites and coverage. To identify essential and non-essential genes, the insertion index was calculated for each gene by dividing the number of unique insertions in the gene by gene length. Observed insertion index values were fitted to a bimodal distribution with a gamma distribution (or an exponential distribution for genes with no observed insertion sites) corresponding to essential and non-essential genes. A $\log_2$ likelihood, and corresponding $P$ values, of each gene belonging to essential or non-essential sets was calculated using R software. To compare the fitness of individual mutants in input and output populations, reads were normalised and tested for differential base means by calculating $\log_2$-fold changes and corresponding $P$ values at a false discovery rate of 0.1 using DESeq with R software. Raw read data for all transposon insertions have been deposited in the European Nucleotide Archive (ENA); accession numbers are as follows: ERR2235345 and ERR2235346 for identification of essential genes for replicates 1 and 2; ERR2235567 for input population; ERR2235568 for output population of rat MSI genes; ERR2235569 for output population of serum-exposed E. coli A192PP; ERR2235570 for output population of bacteria exposed to heat-inactivated serum.

**Colonisation and infection of neonatal rats.** Timed-birth Wistar rat pup litters (usually $n = 12$) were purchased from Harlan UK, delivered at P2 and colonized on the same day. Pups were retained throughout each experiment with the natural mothers in a single dedicated cage under optimal conditions (19-21°C, 45-55% humidity, 15-20 changes of air/h, 12 h light/dark cycle) and were returned to the mother immediately after colonization. Mothers...
had unrestricted access to standard rat chow and water. The procedure has been described in detail (84). In brief, all members of P2 rat pup litters were fed 20μl of mid-logarithmic-phase *E. coli* (2-6 x 10⁶ CFU unless otherwise stated) from an Eppendorf micropipette. GI colonization was confirmed by culture of perianal swabs on MacConkey agar and bacteremia detected by MacConkey agar culture of blood taken *post mortem*. Disease progression was monitored by daily evaluation of symptoms of systemic infection and neonates culled by decapitation and recorded as dead once a threshold had been reached: pups were regularly examined for skin color, agility, agitation after abdominal pressure, presence of a milk line, temperature, weight and behaviour in relation to the mother. Neonates were culled immediately when abnormalities for three of these criteria were evident. After sacrifice, GI tissues were excised aseptically without washing, colon separated and the SI segmented into 2 cm portions representing proximal, middle and distal small intestinal tissue. Tissues were then transferred to ice-cold phosphate-buffered saline, and homogenized. Bacteria were quantified by serial dilution culture on MacConkey agar supplemented with 25 μg/ml kanamycin as appropriate. The presence of *E. coli* K1 was confirmed with phage K1E: 20 lactose-fermenting colonies were streaked onto MH agar, 10 μl of ~10⁹ PFU/ml phage suspension dropped on each streak and the plates incubated overnight. *E. coli* K1 bacteria were quantified by multiplying total CFU by the proportion of K1E susceptible colonies. In all cases at least 19 colonies were susceptible to the K1 phage; *E. coli* K1 was never found in samples from non-colonized pups.

**Susceptibility to human serum.** Serum was obtained from healthy volunteers and used immediately. Bacteria were grown to late logarithmic phase in LB broth in an orbital incubator (minimum 200 orbits/min), 500 μl culture removed, washed twice with gelatin-
Veronal buffered saline plus magnesium and calcium ions (pH 7.35) (GVB++), and suspended in an equal volume of GVB++. Fresh human serum was diluted 1:3 in GVB++ and pre-warmed to 37°C. Bacterial suspensions and serum solutions were mixed 1:2 to give a final concentration of ~10^7 CFU/ml and incubated at 37°C for 3 h in a total volume of 125 µl containing 22% serum. Surviving E. coli were quantified by serial dilution and overnight incubation on LB agar. Pre-warmed, heat-inactivated (56°C, 30 min) serum served as control.

ACKNOWLEDGMENTS

This work was supported by research grant MR/K018396/1 from the Medical Research Council. The National Institute for Health Research University College London Hospitals Biomedical Research Centre provided infrastructural support.
REFERENCES


the genotoxin colibactin and the siderophore yersiniabactin by *Escherichia coli*. J Infect Dis 214:916-924.


44. Ho TD, Davis BM, Ritchie JM, Waldor MK. 2008. Type 2 secretion promotes enterohemorrhagic Escherichia coli adherence and intestinal colonization. Infect Immun 76:1858-1865.


FIG 1. A high-density transposon library for identification of genes essential for in vitro growth of *E. coli* K1 A192PP. (A) Distribution of Tn5 insertions along the *E. coli* K1 A192PP genome. The number of sequence reads mapped to each single genomic location are plotted to show representation of the entire genome. (B) Insertion index values for two biological replicates are strongly correlated. (C) Density plot showing the frequency of insertion index values for all genes. A biomodal distribution is evident, with the left peak representing “essential” genes in which Tn5 insertion is lethal for growth on selective Luria-Bertani agar; the left peak represents “non-essential” genes into which Tn5 inserted without induction of lethality. Green and red lines indicate gamma distributions used to estimate likelihood ratios and *P* values. (D) Tn5 insertion site reads plotted to a 9 Kb region of the *E. coli* A192PP genome. The height of each line on the y axis indicates the number of reads at each Tn5 insertion site. The genes *lytB* and *dapB* possess no insertion sites, indicating they are putative essential genes.

FIG 2. Essential *E. coli* A192PP genes in each selected KEGG (Kyoto Encyclopedia of Genes and Genomes) functional orthologs (KO). Gene frequencies (light grey; expressed as % of essential coding DNA sequences [CDSs] for each category) are compared to their frequency within the whole genome (dark grey). KO (x/y) where y is the number of CDSs in the whole genome and x is the number of identified essential genes.

FIG 3. Culture of the *E. coli* K1 A192PP-Tn5 library results in loss of population diversity and enrichment of non-encapsulated mutants. (A) Changes in the proportion of *E. coli* encapsulated and non-encapsulated A192PP bacteria during culture of the *E. coli* A192PP-Tn5 library in LB media at 37°C (200 orbits/min) (*n* = 3; ±1SD; Student’s *t*, *p* < 0.01). CFU of
encapsulated and non-encapsulated bacteria were determined from the proportion of bacteria susceptible to the K1E bacteriophage. (B) Survival of P2 rats colonized with *E. coli* K1 A192PP, the uncultured *E. coli* A192PP-Tn5 library and the cultured (LB broth; 8 h; 37°C) *E. coli* A192PP-Tn5 library. Pups (*n* = 12 for each group) were colonized with 2-4 x 10⁶ CFU by the oral route. Log-rank [Mantel-Cox] to compare survival of the cultured library with wildtype strain and the uncultured library: ns, non-significant, *P* < 0.05, **P** < 0.01.

**FIG 4.** Identification using a high-density transposon library of genes promoting GI colonization of *E. coli* A192PP in the neonatal rat. (A) Colonization of PSI, MSI, DSI and colon after oral administration of 2-6x10⁶ CFU *E. coli* K1 A192PP to P2 pups. (B) Log₂-fold change and average Tn5 insertion site read abundance of each gene after MSI colonization of P2 rat pups (*n* = 4) over a 4 h stabilization period expressed as MA-plot. An inoculum containing 2 x 10⁴ unique *E. coli* K1 A192PP-Tn5 mutants was prepared and 1 x 10⁹ CFU administered orally. *E. coli* colonies (2 x 10⁵) were recovered from the inoculum (input pool) and from MSI homogenates (output pool) by culture on to LB agar containing 50 µg/ml kanamycin. Red data points represent Tn5 insertion sites determined as differentially expressed in the output pool compared to the input pool using a negative binomial test with a false discovery rate of 0.1. (C) Mutations in 167 genes significantly decreased fitness for colonization of the MSI and encoded proteins with a range of functions. (D) Colonization of P2 rat intestine by *E. coli* K1 A192PP and single gene mutants. Bacteria (2-4 x 10⁶) were administered orally to P2 rats (*n* = 12/group). Pups were sacrificed and the *E. coli* K1 burden in intestinal sections enumerated 24 h after initiation of colonization. Parent and mutant strain CFU values were compared using Student’s *t*-test: *P* < 0.05, **P** < 0.01. (E) Survival of P2 rats colonized with *E. coli* K1 A192PP and single gene mutants. Bacteria (2-4 x10⁶) were administered orally to
FIG 5. Bottleneck to systemic infection in the neonatal rat. (A) Survival of rats colonized at P2 by oral administration of E. coli K1 A192PP or A192PPΔlacZ::kan. n = 12 pups for both groups. Log-rank [Mantel-Cox] test: ns, non-significant, * P < 0.05, ** P < 0.01. (B) Competitive indices of intestinal colonization and gut-to-blood transit of E. coli K1 A192PP and A192PPΔlacZ::kan. A 1:1 mixture of E. coli K1 A192PP and A192PPΔlacZ::kan (total 2–4 x10⁶ CFU) was administered orally to P2 pups. After 24 h, ratios of A192PP and A192PPΔlacZ::kan were enumerated in segmented GI tissues and in the blood as indicated using selective media. Animals in which only E. coli K1 A192PP or A192PPΔlacZ::kan were detected in the blood are coloured red and blue respectively, indicating the existence of a bottleneck to infection. (C) Loss of diversity of E. coli K1 A192PP-Tn5 populations recovered from the blood (red) and brain (blue) following translocation from the GI tract (black).

FIG 6. Identification using a high-density transposon library of genes contributing to the complement resistance of E. coli A192PP. (A) Survival of E. coli A192PP and E. coli K12 strain MG1655 in 22% pooled human serum. The latter was used a serum susceptible control; n = 3, error bars represent range of values. (B) Log₂-fold change and average Tn5 insertion site read abundance of each gene after incubation of 1 x 10⁶ CFU containing 2 x 10⁴ unique E. coli K1 A192PP-Tn5 mutants in 22% pooled human serum for 3 h at 37°C. Colonies (2 x 10⁵) were obtained by culture of diluted aliquots on LB agar containing 50 µg/ml kanamycin. The inoculum served as the input pool. Red data points represent Tn5 insertion sites determined as differentially expressed in the output pool compared to the input pool using a negative binomial test with a false discovery rate of 0.1. (C) Survival of 1 x10⁶ E. coli K1 A192PP and
single gene mutants in 22% normal human and heat-inactivated (56°C; 30 min) serum. Final volume of the reaction mixture was 1.5 ml; n = 3, error bars represent range of values. Complementation with the functional gene restored resistance in all cases. (D) Survival of P2 rats colonized with E. coli K1 A192PP and single gene mutants. Bacteria (2-4 x10⁶) were administered orally to P2 rats (n = 12/group). Log-rank [Mantel-Cox] test: ns, non-significant, * P < 0.05, ** P < 0.01.
### TABLE 1. Strains used in this study

<table>
<thead>
<tr>
<th>Strain</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em> K1</td>
<td></td>
</tr>
<tr>
<td>A192PP</td>
<td>018:K1:H7; virulent in neonatal rat model of infection</td>
</tr>
<tr>
<td>A192PPΔlacZ::kan</td>
<td>LacZ mutant of A192PP; Kan&lt;sup&gt;r&lt;/sup&gt;</td>
</tr>
<tr>
<td>A192PPΔneuC::kan</td>
<td>NeuC mutant of A192PP; Kan&lt;sup&gt;r&lt;/sup&gt;</td>
</tr>
<tr>
<td>A192PPΔneuC::kan + pUC19.neuC</td>
<td>Complemented A192PPΔneuC::kan; Kan&lt;sup&gt;r&lt;/sup&gt;Amp&lt;sup&gt;r&lt;/sup&gt;</td>
</tr>
<tr>
<td>A192PPΔrfaH::kan</td>
<td>RfaH mutant of A192PP; Kan&lt;sup&gt;r&lt;/sup&gt;</td>
</tr>
<tr>
<td>A192PPΔrfaH::kan + pUC19.rfaH</td>
<td>Complemented A192PPΔrfaH::kan; Kan&lt;sup&gt;r&lt;/sup&gt;Amp&lt;sup&gt;r&lt;/sup&gt;</td>
</tr>
<tr>
<td>A192PPΔtraL::kan</td>
<td>TraL mutant of A192PP; Kan&lt;sup&gt;r&lt;/sup&gt;</td>
</tr>
<tr>
<td>A192PPΔtraL::kan + pUC19.traL</td>
<td>Complemented A192PPΔtraL::kan; Kan&lt;sup&gt;r&lt;/sup&gt;Amp&lt;sup&gt;r&lt;/sup&gt;</td>
</tr>
<tr>
<td>A192PPΔvasL::kan</td>
<td>VasL mutant of A192PP; Kan&lt;sup&gt;r&lt;/sup&gt;</td>
</tr>
<tr>
<td>A192PPΔwaaW::kan</td>
<td>WaaW mutant of A192PP; Kan&lt;sup&gt;r&lt;/sup&gt;</td>
</tr>
<tr>
<td>A192PPΔwaaW::kan + pUC19.waaW</td>
<td>Complemented A192PPΔwaaW::kan; Kan&lt;sup&gt;r&lt;/sup&gt;Amp&lt;sup&gt;r&lt;/sup&gt;</td>
</tr>
<tr>
<td>A192PPΔyaeQ::kan</td>
<td>YaeQ mutant of A192PP; Kan&lt;sup&gt;r&lt;/sup&gt;</td>
</tr>
<tr>
<td>A192PPΔ0678::kan</td>
<td>0678 mutant of A192PP; Kan&lt;sup&gt;r&lt;/sup&gt;</td>
</tr>
<tr>
<td>A192PPΔ3010::kan</td>
<td>3010 mutant of A192PP; Kan&lt;sup&gt;r&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>E. coli</em> K12</td>
<td></td>
</tr>
<tr>
<td>MG1655</td>
<td>F- lambda- ilvG- rfb-50 rph-1</td>
</tr>
<tr>
<td>MG1655ΔlacZ::kan</td>
<td>LacZ mutant of MG1655; Kan&lt;sup&gt;r&lt;/sup&gt;</td>
</tr>
</tbody>
</table>
A

![Graph showing survival rates over days after oral K1 dose]

B

![Scatter plot showing competitive index across different samples]

C

![Genome sequence analysis showing normalized reads in MSI, Blood, and Brain samples]

Days after oral K1 dose

Competitive Index

PSI, MSI, DSI, Colon, Blood

Normalized Reads

Kb