





A Novel Protective Vaccine Antigen from the Core *Escherichia coli* Genome

Danilo G. Moriel,^{a,b*} Lendl Tan,^{a,b} Kelvin G. K. Goh,^{a,b}  Minh-Duy Phan,^{a,b}
Deepak S. Ipe,^c Alvin W. Lo,^{a,b} Kate M. Peters,^{a,b}  Glen C. Ulett,^c
 Scott A. Beatson,^{a,b}  Mark A. Schembri^{a,b}

School of Chemistry and Molecular Biosciences, the University of Queensland, Queensland, Brisbane, Australia^a; Australian Infectious Diseases Research Centre, the University of Queensland, Queensland, Brisbane, Australia^b; School of Medical Sciences and Menzies Health Institute Queensland, Griffith University, Gold Coast, Queensland, Australia^c

ABSTRACT *Escherichia coli* is a versatile pathogen capable of causing intestinal and extraintestinal infections that result in a huge burden of global human disease. The diversity of *E. coli* is reflected by its multiple different pathotypes and mosaic genome composition. *E. coli* strains are also a major driver of antibiotic resistance, emphasizing the urgent need for new treatment and prevention measures. Here, we used a large data set comprising 1,700 draft and complete genomes to define the core and accessory genome of *E. coli* and demonstrated the overlapping relationship between strains from different pathotypes. In combination with proteomic investigation, this analysis revealed core genes that encode surface-exposed or secreted proteins that represent potential broad-coverage vaccine antigens. One of these antigens, YncE, was characterized as a conserved immunogenic antigen able to protect against acute systemic infection in mice after vaccination. Overall, this work provides a genomic blueprint for future analyses of conserved and accessory *E. coli* genes. The work also identified YncE as a novel antigen that could be exploited in the development of a vaccine against all pathogenic *E. coli* strains—an important direction given the high global incidence of infections caused by multidrug-resistant strains for which there are few effective antibiotics.

IMPORTANCE *E. coli* is a multifaceted pathogen of major significance to global human health and an important contributor to increasing antibiotic resistance. Given the paucity of therapies still effective against multidrug-resistant pathogenic *E. coli* strains, novel treatment and prevention strategies are urgently required. In this study, we defined the core and accessory components of the *E. coli* genome by examining a large collection of draft and completely sequenced strains available from public databases. This data set was mined by employing a reverse-vaccinology approach in combination with proteomics to identify putative broadly protective vaccine antigens. One such antigen was identified that was highly immunogenic and induced protection in a mouse model of bacteremia. Overall, our study provides a genomic and proteomic framework for the selection of novel vaccine antigens that could mediate broad protection against pathogenic *E. coli*.

KEYWORDS: *Escherichia coli*, vaccines, virulence factors

Escherichia coli is a bacterium that exists in many different guises. On one hand, it is a common commensal organism of the human small intestine. On the other, it is a rapidly evolving pathogen that is able to acquire and combine different genetic elements into novel and complex gene repertoires. The latter has led to the evolution of *E. coli* as a multifaceted pathogen, highlighted by aggressive disease outbreaks (1) and the emergence of multidrug-resistant (MDR) lineages (2, 3).

Received 26 October 2016 Accepted 27 October 2016 Published 23 November 2016

Citation Moriel DG, Tan L, Goh KGK, Phan M-D, Ipe DS, Lo AW, Peters KM, Ulett GC, Beatson SA, Schembri MA. 2016. A novel protective vaccine antigen from the core *Escherichia coli* genome. *mSphere* 1(6): e00326-16. doi:10.1128/mSphere.00326-16.

Editor Melanie Blokesch, Swiss Federal Institute of Technology Lausanne

Copyright © 2016 Moriel et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Mark A. Schembri, m.schembri@uq.edu.au.

*Present address: Danilo G. Moriel, GSK Vaccines Institute for Global Health S.r.l., Siena, Italy.

TABLE 1 Distribution of strains in EcoDS according to source, disease, year, and place of isolation

Source	%	Disease ^a	%	Yr	%	Place	%
Bat	0.1	Asymptomatic	0.8	<1940s	0.1	Africa	0.8
Buffalo	0.1	Bacteremia	13.1	1950s	0.5	Asia	15.6
Cow	4.5	Bacteriuria	0.1	1960s	0.6	Europe	7.3
Dog	0.2	Crohn's disease	0.2	1970s	0.7	North America	19.4
Environment	0.9	Diarrhea	6.7	1980s	2.2	South America	1.5
Fish	0.1	HUS	1.1	1990s	5.5	Oceania	0.8
Food	1.0	Mastitis	0.4	2000s	12.2	Unknown	54.6
Goat	0.1	Meningitis	0.2	2010s	9.6		
Horse	0.1	Omphalitis	0.1	Unknown	68.6		
Human	43.9	Peritonitis	0.1				
Marsupial	0.9	RTI	0.1				
Mouse	0.1	Septicemia	0.2				
Pig	1.1	UTI	2.1				
Poultry	0.8	Unknown	74.8				
Rabbit	0.2						
Reptile	0.1						
Sheep	0.1						
Wild bird	0.6						
Unknown	45.1						

^aHUS, hemolytic-uremic syndrome; RTI, respiratory tract infections; UTI, urinary tract infections.

E. coli can be classified into different pathotypes according to a common set of virulence factors and specific clinical manifestations (4). Despite these phenotypic associations, strains from a single pathotype are not restricted to one phylogroup; such strains can share the same genomic profile with other pathotypes (5) and be distributed over the entire span of the *E. coli* phylogenetic diversity (6, 7). These observations indicate a common evolutionary origin and divergence into different pathotypes as a result of the independent acquisition of specific virulence genes via multiple events of horizontal gene transfer (8).

The 2011 *E. coli* O104:H4 German outbreak provided a new perspective to our understanding of evolution and genome plasticity in the species. The outbreak strain had acquired key virulence genes from two different *E. coli* pathotypes (enteroaggregative *E. coli* [EAEC] and enterohemorrhagic *E. coli* [EHEC]) and, combined with genes encoding resistance to antibiotics, emerged as a highly virulent lineage that infected nearly 4,000 people and caused 54 deaths (9). Since this outbreak, it has been proposed that targeting of accessory components encoded by the *E. coli* genome may be insufficient to prevent the emergence of new pathogenic lineages, and broader strategies directed against conserved features of all strains may be more effective (10).

E. coli represents one of the most sequenced microorganisms in public databases, with draft and complete genome sequences available for strains isolated from different hosts, disease associations, and geographic locations. In this study, we took advantage of this large genome sequence resource to define core and accessory *E. coli* genes. Furthermore, we used the resultant data to better understand the structural phylogeny of *E. coli* and identified YncE as a highly conserved vaccine antigen that is protective against acute *E. coli* bacteremia.

RESULTS

EcoDS: a large data set of *E. coli* genome sequenced strains. An *E. coli* data set (EcoDS) was generated from 1,700 genome sequences available on the NCBI public database. EcoDS contained 62 complete and 1,638 draft genome sequences and represented a highly diverse collection of *E. coli* strains (Table 1; see Fig. S1 in the supplemental material). Information regarding the source, disease, origin, or year of isolation was available for approximately 69% ($n = 1,174$) of strains. Analysis of this strain subset revealed the most common sources as human ($n = 747$) or cattle ($n = 76$) and the most common disease associations as bacteremia ($n = 222$) or diarrhea ($n = 114$). The majority of strains originated from North America ($n = 327$) or Asia ($n = 265$).

(Table 1). Overall, EcoDS is represented by commensal *E. coli* strains as well as strains from all known *E. coli* pathotypes. This includes the intestinal pathotypes EAEC, EHEC, enteroinvasive *E. coli* (EIEC), adherent-invasive *E. coli* (AIEC), enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), and the recently emerged enteroaggregative and hemorrhagic *E. coli* (EAHEC) pathotype, as well as extraintestinal pathotypes comprising avian-pathogenic *E. coli* (APEC), neonatal meningitis *E. coli* (NMEC), and uropathogenic *E. coli* (UPEC).

Phylogenetic and pathotype relationship of *E. coli* strains in EcoDS. Phylogenetic group determination and multilocus sequence typing (MLST) were performed to characterize the relationship of all of the strains in EcoDS. Using this combined analysis, a phylogenetic tree was constructed based on 435 unique MLSTs identified in EcoDS (Fig. 1). This strategy confirmed the strong correlation between *E. coli* STs and the established multilocus enzyme electrophoresis (MLEE)-based *E. coli* phylogeny (11) and described a comprehensive distribution of STs within the major phylogroups in EcoDS (Fig. 1).

To establish a correlation between the *E. coli* phylogeny and pathotypes, the major phylogroups were analyzed in greater detail by a comparison of multiple factors, including year of isolation, origin, source, and disease association (Table 1). No clear geographic, host, or disease-specific component correlated with this clonal phylogenetic framework, highlighting a limitation for analysis of large collections with incomplete clinical sampling information (12).

EcoDS contained a highly diverse combination of STs and pathotypes (Fig. 1). The composition of phylogroup A includes K-12 (MG1655, W3110, MC4100, MDS42, c321.deltaA, BW2952, and DH10B) and B [BL21(DE3) and BL21-Gold(DE3)(pLysS) AG] strains, as well as most of the complete ETEC genome sequences and the commensal strain HS. The K-12 and UMN18 strains belong to ST10, which is also represented by strains from UPEC (ATCC 23506), EAEC (C43/90), and APEC (S17) pathotypes. Phylogroup AxB1 is a hybrid group that comprehends strains from phylogroups A and B1. It contains the ST678 EAHEC strains involved in the 2011 German outbreak (1, 13–15) and strains from progenitor EHEC and EAEC STs that contributed to the emergence of this highly virulent lineage (e.g., the EAEC strain 55989 and the EHEC strains 2009EL-2050 and 2009EL-2071). Interestingly, phylogroup AxB1 is also represented by ST675, recently described as containing EHEC strains involved in UTI cases (16). Moreover, this group also contains nonpathogenic strains and strains from APEC and ETEC pathotypes. Phylogroup B2 contains a large number of strains from ST73 ($n = 116$) and the recently emerged and globally disseminated MDR ST131 clonal lineage ($n = 38$) (17, 18). This phylogroup predominantly contains strains associated with extraintestinal human infections, although strains from AIEC and EPEC pathotypes are also present. Phylogroup D is represented by the clonal lineages including subgroup E, the previously described subgroup F (6), and the recently identified cryptic lineages referred to as C-I to C-V (19, 20). Phylogroup E is the least diverse group and is a clonal lineage predominantly represented by EHEC and EPEC strains belonging to ST11 (235/272 strains) and from O157 and O55 serogroups. Phylogroup D also comprehends the subgroup F, which is represented by the environmental isolate SMS-3-5, the NMEC strain ϵ E10, and the UPEC strain IAI39. The cryptic lineages C-I to C-V include EHEC, ETEC, commensal strains, and the environmental isolates previously described (19, 20). Moreover, phylogroup D is also represented by strains from ST69 ($n = 80$), which includes the UPEC reference strain UMN026 and EAEC strain 42. Together, our analysis shows that strains from the same phylogroup, and even from the same ST, can be involved in different clinical manifestations and belong to different pathotypes.

Determination of the *E. coli* core and accessory genome. The majority of genomes within EcoDS comprised draft genome sequences, and we predicted that they would vary significantly in quality and coverage. Thus, we devised a strategy based on the prevalence of *E. coli* essential genes to remove genome sequences with poor or low coverage. First, we used two recent studies (21, 22) to define a set of 362 *E. coli*

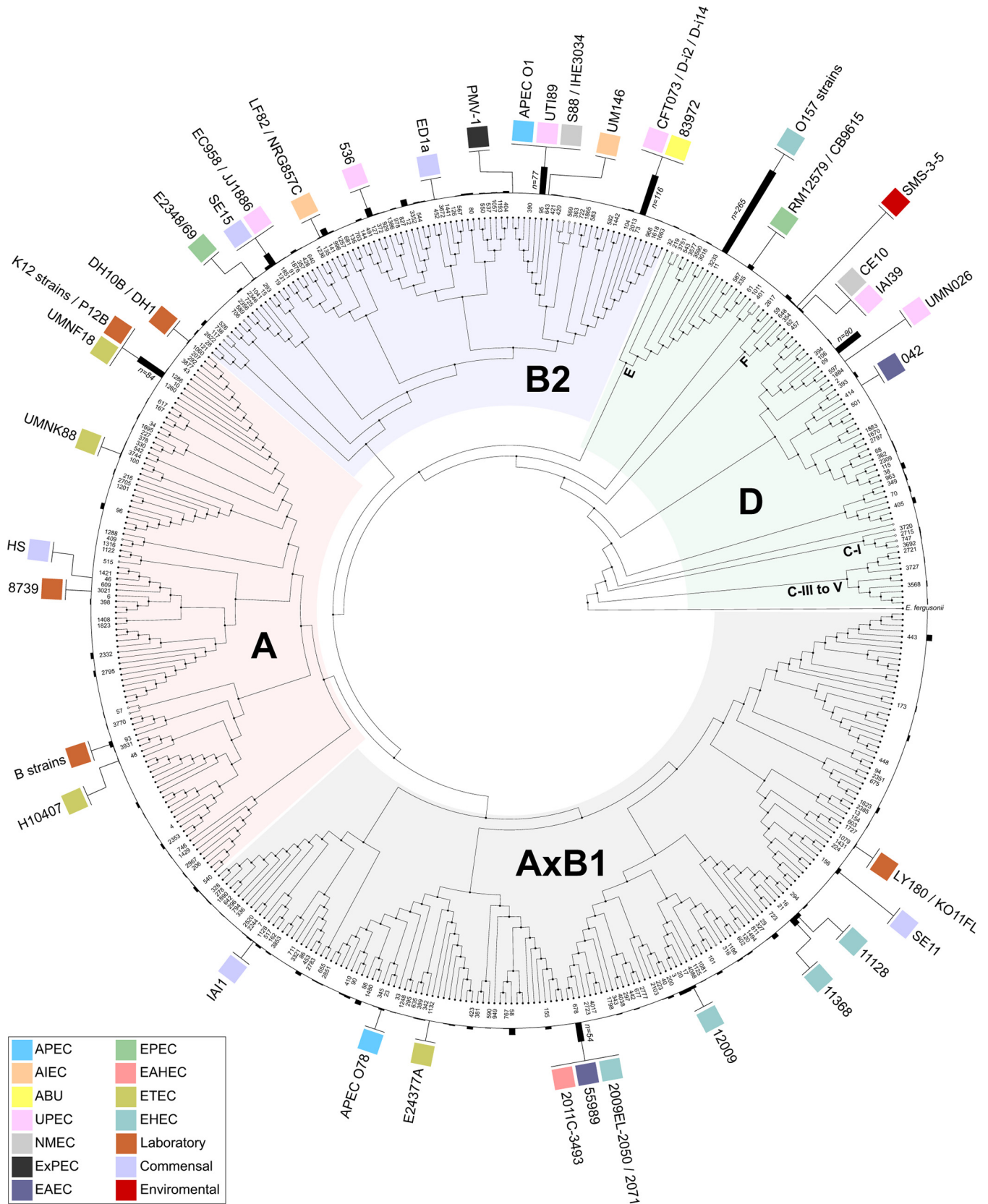


FIG 1 Phylogeny of *E. coli* strains in EcoDS. Phylogenetic tree demonstrating the relationship of MLSTs in EcoDS. Major phylogroups A, B1, B2, and D, as well as minor phylogroups E, F, C-I, and C-III to -V, are indicated. Completely sequenced strains are indicated according to their pathotype association (see the inset for the color-coded legend): (i) O157 strains, EDL933, Sakai, TW14359, EC4115, and Xuzhou21; (ii) K-12 strains, MC4100, MDS42, W3110, MG1655, c321.deltaA, and BW2952; and (iii) B strains, REL606, BL21(DE3), and BL21-Gold(DE3)(pLysS) AG. Histograms represent the number of genomes

(Continued)

essential genes (see Data Set S1 in the supplemental material). We then screened for the prevalence of these essential genes in the 62 complete genomes, which should represent the best quality genome sequences in EcoDS. In total, 318 genes were present in all 62 completely sequenced strains. These 318 genes were therefore used to filter the 1,700 genomes in EcoDS, and our analysis revealed a prevalence of $99.64\% \pm 3.0\%$ essential genes per strain (which allows a mean tolerance of one missing essential gene per strain). Therefore, strains missing more than one essential gene were discarded ($n = 144$ [see Data Set S2 in the supplemental material]), leaving a total of 1,556 strains in EcoDS. The least prevalent essential gene in EcoDS was *b3119* (*tdcR*), which was present in 99% ($n = 1,541$) of strains (see Data Set S3 in the supplemental material).

We used the 1,556 genome sequences in EcoDS to define the conserved set of *E. coli* genes. To enable this analysis, we used the well-characterized and best-annotated K-12 strain MG1655 as a reference (23), and based on the essential gene data, a cutoff value for gene prevalence was set at 99%. Pairwise comparison of the 4,319 open reading frames (ORFs) defined in MG1655 with the genome sequence of the 1,556 strains in EcoDS led to the identification of 3,042 genes present in more than 99% of strains (protein identity of $>75\%$ over a 75% sequence overlap), of which 1,037 genes were present in 100% of the strains. These 3,042 genes define a conserved subset of *E. coli* genes in the 1,556 strains that make up EcoDS (Fig. 2; see Data Set S4 in the supplemental material).

In order to define the *E. coli* accessory genome, 298,563 annotated ORFs (which comprised 290,776 chromosomal ORFs and 7,787 plasmid ORFs) from the 62 completely sequenced strains were compared to each other. This pairwise comparison identified 12,722 unique ORFs based on $>75\%$ protein identity that displayed a prevalence of $<99\%$ in EcoDS (10,513 chromosomal and 2,209 plasmid [see Data Set S5 in the supplemental material]). While we accept that our arbitrary cutoff of $>75\%$ protein identity may result in the overlooking of some biological variation, as previously described in *Shigella* (24) and EPEC (25), our analysis nevertheless reflects the enormous genetic diversity that exists in the pan-*E. coli* genome. We refer to this gene set as the accessory *E. coli* genome. We note that no genes absent in MG1655 exhibited $>99\%$ prevalence in EcoDS, thus providing support to our overall determination of the core genome (1,037 genes) and the conserved subset of *E. coli* genes (3,042 genes).

Identification of novel vaccine targets. The definition of a conserved set of 3,042 *E. coli* genes provides a framework for the development of new tools in epidemiology, diagnostics, and vaccine antigen discovery. In order to evaluate the expression of core components as potential antigens that could be targeted for vaccine development, we examined the proteome of strain EC958, which belongs to a clinically relevant and globally disseminated MDR sequence type (ST131). Proteomic analysis of outer membrane vesicles (OMVs) induced from EC958 led to the identification of 115 proteins (see Data Set S6 in the supplemental material). Among them, 23 were predicted to have an extracellular ($n = 1$), outer membrane ($n = 15$), or unknown ($n = 7$) subcellular localization. To refine this list, we focused our attention on proteins encoded by genes that were prevalent in more than 99% of EcoDS. This led to a panel of 17 potential surface-exposed antigens, which was even further reduced by removing outer membrane proteins predicted to have a transmembrane β -barrel structure based on analysis by PHYRE2 (i.e., OmpA, OmpC, OmpF, OmpX, MipA, Fiu, FepA, Tsx, and CirA) and proteins with an unknown subcellular localization and no predicted signal sequence based on LipoP (Wrba, SodA, TrmL, and LysC). This left a final list of four potential surface-associated proteins, namely, BamC, OsmE, SlyB, and YncE.

YncE is a highly immunogenic and protective antigen. To examine these four proteins further, each respective gene was PCR amplified and cloned in frame with an

Figure Legend Continued

in the respective ST within EcoDS. STs represented by more than 50 genomes are indicated. Open circles represent STs that belong to a different phylogroup.

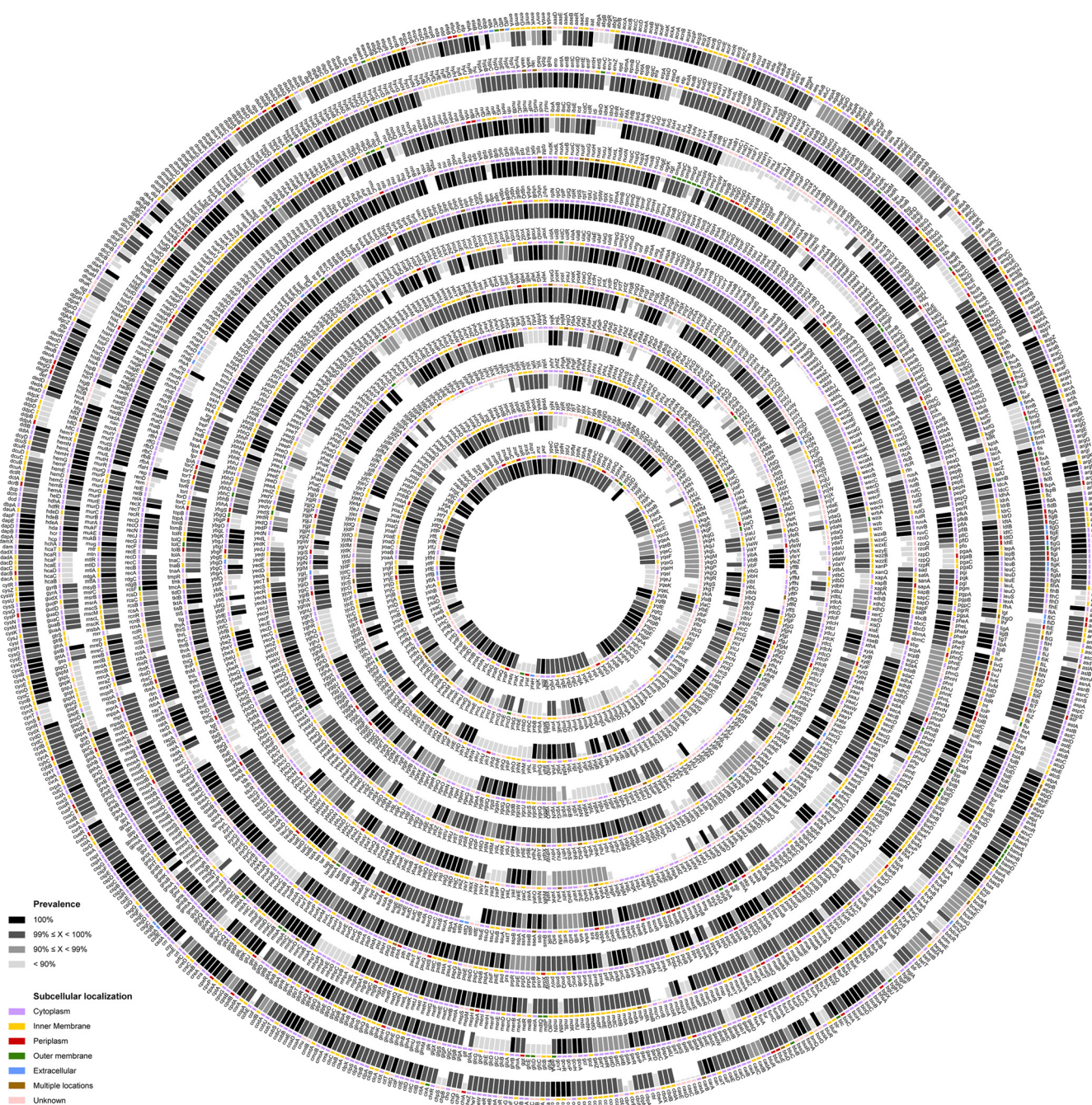


FIG 2 The *E. coli* core genome. The distribution of MG1655 genes among strains in EcoDS is represented by histograms and color coded according to prevalence and subcellular localization (predicted by PSORTb). Only the 4,319 unique genes in the MG1655 genome are shown.

N-terminal 6×His tag sequence. Expression studies using these constructs revealed only the BamC and YncE proteins were produced as soluble recombinant proteins (see Fig. S2A in the supplemental material). Therefore, these two proteins were purified and tested for immunogenicity using plasma obtained from convalescent urosepsis patients and plasma from an age- and sex-matched healthy control group. As a positive control and correlate of immunogenicity, we also included SsIE (ecok1_3385) and EsiB (c5321) in our analysis, both of which have previously been shown to be protective against extraintestinal pathogenic *E. coli* (ExPEC) infection in a mouse sepsis model (26). Both SsIE ($P < 0.0001$) and EsiB ($P < 0.01$) showed higher reactivity to the plasma from urosepsis patients compared to healthy individuals (Fig. 3A). Among the targets

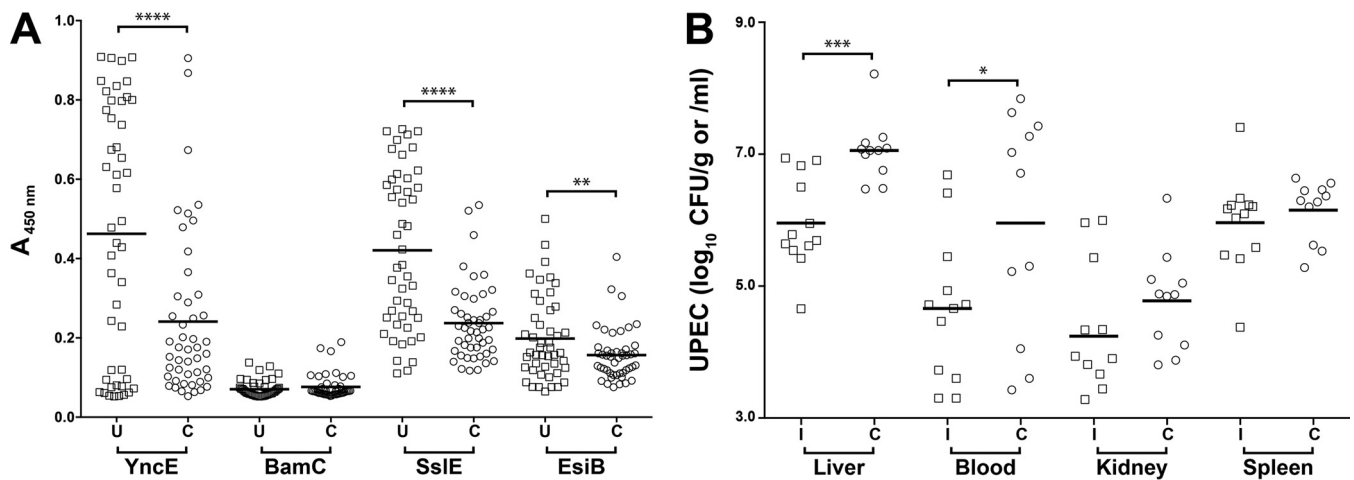


FIG 3 Immunoreactivity of plasma from urosepsis patients to *E. coli* vaccine antigens and bacterial burden following intravenous challenge in mice vaccinated with YncE. (A) Blood plasma was collected from 47 urosepsis patients (U) at least 4 days post-admittance to the hospital. IgG-specific antibody levels were compared to those from 47 healthy volunteers with no recent history of UTI (C). (B) Level of infection in immunized (I) and control (C) groups of mice at 24 h following intravenous challenge with UPEC strain CFT073. Symbols represent individual mice, and bars show the medians. The limit of detection was 200 CFU/g or CFU/ml. Statistically significant *P* values are indicated: ****, $P < 0.0001$; ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$.

identified in this study, BamC showed no significant reactivity with plasma from urosepsis patients. However, in contrast, YncE was strongly reactive with plasma from urosepsis patients compared to healthy individuals ($P < 0.0001$), suggesting it is expressed during human infection.

We investigated the ability of YncE to elicit a protective immune response against acute systemic *E. coli* infection using an established murine model of bacteremia (27). YncE was highly immunogenic, inducing a strong IgG response (Fig. S2B), and mice immunized with YncE were significantly protected against infection, as evidenced by lower blood and liver *E. coli* loads following systemic challenge (Fig. 3B). Taken together, these data identify YncE as a novel, highly conserved and strongly immunogenic *E. coli* antigen that is able to provide protection against acute systemic infection when administered by vaccination.

YncE is broadly expressed and secreted by different *E. coli* pathotypes. The *yncE* gene was highly conserved among strains in EcoDS ($n = 1,550/1,556$ [99.6%]). Strains lacking the *yncE* gene were not related by ST or phylogroup: DEC13A, ST296/B1; DEC13B, ST296/B1; DEC13D, ST278/B1; IAI39, ST24/D; N1, ST515/B1; and UMEA 3687-1, ST73/B2. To verify the expression of YncE by strains belonging to different pathotypes and phylogroups, Western blot analysis was performed using rabbit polyclonal antibodies generated against recombinant YncE (Fig. 4A). It was observed that YncE is broadly expressed by *E. coli* strains from different pathotypes, including ETEC (strain H10407), EHEC (strain Sakai), EAEC (strain 42), and UPEC (strains NDM10, EC958, and CFT073). Moreover, Western blot analysis of strains representing the most predominant *E. coli* phylogroups and subgroups (A, AxB1, E, D, F, and B2) determined in this study revealed the presence of YncE in the supernatant after overnight growth (Fig. 4A). YncE was also overexpressed in an *E. coli fur* mutant during *in vitro* growth, demonstrating a possible role for iron in its regulatory control (Fig. 4B).

DISCUSSION

Current vaccination strategies against *E. coli* have focused on individual pathotypes and targeted major virulence determinants. Examples include colonization factor antigens and heat-labile toxin from ETEC (28), components and effectors of the type III secretion system from EPEC (29), Shiga toxin from EHEC (30), and fimbrial adhesins and siderophore receptors from UPEC (31). In this study, we used an essential gene strategy to generate a curated data set of *E. coli* genomes (EcoDS) and used this diverse collection (representing 435 STs) to define core and accessory elements of the pan-*E. coli* genome.

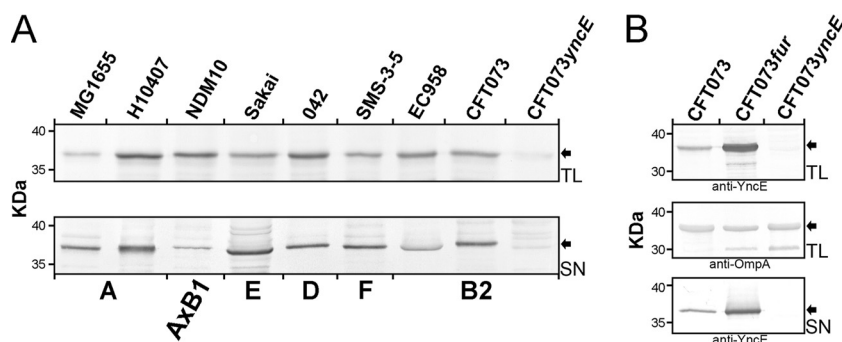


FIG 4 Western blot analysis of total cell lysate (TL) and supernatant (SN) fractions prepared from *E. coli* strains representing different phylogroups. (A) The YncE protein was detected in all TL and SN samples. The specificity of the antiserum was confirmed by the absence of a cross-reacting band in the CFT073 *yncE* mutant. (B) Overexpression of YncE in a *fur* knockout mutant. OmpA was used as an expression control.

This information, together with proteomic data, led to the identification and validation of YncE as a highly conserved and protective *E. coli* vaccine antigen.

We observed that the clustering of different strains according to the seven house-keeping MLST genes was consistent with a modern framework composed of phylogroups A, AxB1, B2, and D (6). Moreover, we observed that single subpopulations comprehended strains involved in multiple diseases, supporting previous studies that demonstrated strains from the same pathotype can be distributed over the entire span of phylogenetic diversity and are not restricted to one specific phylogroup (5–7). Therefore, despite the high diversity, complexity, and low prevalence of the accessory elements in EcoDS, *E. coli* strains possess a defined set of highly prevalent genes irrespective of pathotype, phylogroup, and associated clinical disease.

Comparative analysis of the genomes in EcoDS showed that genes comprising the *E. coli* accessory genome were highly variable in prevalence. The accessory genome is considered a flexible gene pool shaped by mobile genetic elements and represents a major driver of *E. coli* evolution (32). In this study, we observed that the *E. coli* accessory genome comprises a broad set of 12,722 ORFs (prevalence of <99% in EcoDS) and that approximately 90% of the entire accessory genome is present in less than 90% of strains in EcoDS, which poses an enormous challenge for broad therapeutic interventions. Potential *E. coli* universal vaccine targets such as FimH and SsIE, for example, were present in approximately 89% and 70% of the strains in EcoDS, respectively.

A smaller subset of 3,042 genes present in more than 99% of strains in EcoDS was also defined. These genes were used to trace novel vaccinology strategies. In order to identify potential surface-exposed and secreted proteins from the *E. coli* core genome that could represent new vaccine targets, we investigated the OMV-associated proteome of the MDR ST131 strain EC958. Although OMVs are not only represented by surface-exposed proteins, its combination with *in silico* subcellular localization prediction tools can facilitate the identification of exposed targets (33). Among the 115 proteins identified, we selected four proteins for cloning and expression based on literature analysis, predicted subcellular localization, and structural conformation. Among them, YncE showed increased reactivity to a plasma collection of urosepsis convalescent patients, higher than the titers obtained for SsIE, a type II secreted mucinase originally identified from ExPEC (26, 34) that provides broad protection against *E. coli* infection in different animal models (34). BamC, a surface-associated lipoprotein (35), showed no significant reactivity with the urosepsis plasma collection. Moreover, YncE was also shown to be highly immunogenic and decreased the level of infection in a murine model of bacteremia, confirming the immunogenicity of this antigen and its potential as a broad vaccine candidate against *E. coli*.

YncE is a seven-bladed beta-propeller (36) transported by the Sec machinery (37) and associated with binding to single-stranded DNA (38). We have previously shown

that YncE is present in the OMV proteome of a large collection of urosepsis strains (39), and in this study, we demonstrated that YncE is present in the OMV proteome of EC958 (phylogenetic group B2) and in the secretome of strains representing all other phylogenetic groups. Moreover, we confirmed the regulation of YncE by Fur (40), indicating its potential role during infection in iron-limiting environments such as blood and the urinary tract. Taken together, our results indicate that YncE fulfills many prerequisites required for a vaccine candidate: YncE is (i) immunogenic, (ii) highly prevalent, (iii) highly conserved, (iv) soluble, (v) stable, and (vi) expressed during infection.

In conclusion, we have demonstrated the genome complexity and plasticity of *E. coli* and dissected the difficulty associated with targeting the accessory genome for broadly therapeutic interventions. Moreover, we confirmed the close association between different pathogenic and nonpathogenic *E. coli* lineages. We also designed a strategy based on the *E. coli* core genome for the identification of novel potential vaccine targets, which led to the discovery of YncE as an immunogenic and protective antigen. Although we cannot predict the impact of this conserved antigen to the microbiome, recent studies have shown that vaccination with conserved antigens indicate no significant interference with the microbiome (41). Even though YncE is a highly prevalent vaccine antigen, we envisage its use in combination with pathotype-specific antigens rather than in standalone formulations, which would also help to define a target population for clinical testing. Additional work is now required to evaluate the efficacy of YncE immunization in different animal infection studies, alone and in combination with other broadly prevalent or pathotype-specific antigens.

MATERIALS AND METHODS

The *E. coli* data set and MLST analysis. The *E. coli* database was represented by 62 complete and 1,638 draft genomes available on the NCBI public database as of 1 January 2014 (see Data Set S7 in the supplemental material).

Bioinformatic analysis. Sequence comparisons were performed using the FASTA36 package (42). Subcellular localization was predicted by PSORTb 3.0 (43), and signal sequence was predicted by LipoP 1.0 (44). Structural analysis was performed by PHYRE2 (45). The core and accessory genome was determined by amino acid sequence identity using tfastx36 (42) and a cutoff of >75% over a 75% alignment. Strains were classified into the major A, B1, B2, and D phylogroups using an *in silico* triplex analysis of the *chuA*, *yjaA*, and TSPE4.C2 loci (46, 47); analysis was performed by nucleotide sequence identity using FASTA36 (42) and a cutoff of >90% over a 90% alignment. Further *in silico* classification of selected strains into other less common phylogroups was performed using an extension of this scheme (48). MLST analysis was performed using the sequence of seven housekeeping genes as previously described (8). Phylogenetic trees were drawn using MEGA6 (49) using the concatenated sequence of the seven housekeeping genes. Circular representations were drawn using Circos (50).

Bacteria and growth conditions. *E. coli* strains were routinely grown at 37°C on solid or in liquid Luria-Bertani (LB) medium supplemented with the appropriate antibiotics: chloramphenicol (30 µg/ml), kanamycin (50 µg/ml), and ampicillin (100 µg/ml). For the generation of total cell lysates and supernatant fractions, bacteria were inoculated in liquid LB to a starting optical density at 600 nm (OD₆₀₀) equal to 0.050, and cultures were grown overnight at 37°C with shaking (180 rpm). Cells were harvested at 10,000 × *g* for 10 min at 4°C to generate the whole-cell lysate sample. Supernatant fractions were generated as previously described (51). To obtain EDTA-treated heat-induced OMVs, *E. coli* EC958 was grown in minimal M9 medium supplemented with Casamino Acids at 37°C under shaking conditions (180 rpm) and harvested at a final OD₆₀₀ of 0.5. Cells were centrifuged at 10,000 × *g* for 10 min at 4°C, and the pellet was used for OMV heat induction.

Molecular methods, proteomic analyses, and immunoblotting. Genomic DNA was extracted using the Wizard Genomic DNA purification kit (Promega). PCR was performed using Phusion High-Fidelity DNA polymerase (New England Labs) and primers described in Data Set S8 in the supplemental material. PCR products were purified from agarose gel using the QIAquick gel extraction kit (Qiagen) and cloned into pMCSG7 using the ligation-independent cloning (LIC) method (52). DNA was transformed into chemically competent *E. coli* TOP10 cells (Invitrogen) for plasmid propagation. The primers used for cloning are described in Data Set S8. Isolation of plasmid DNA was carried out using the QIAprep spin miniprep kit (Qiagen), and DNA was sequenced using the BigDye Terminator v3.1 cycle sequencing kit (Invitrogen). Correct constructs were used to transform *E. coli* BL21(DE3)(pLysS) (Invitrogen) for protein expression. A CFT073 *yncE* mutant was constructed by λ-Red mediated homologous recombination using primers 5600 and 5601 (Data Set S8) as previously described (53). The formation of OMVs was induced by incubation in 0.01 M EDTA at 56°C, and samples were analyzed by high-performance liquid chromatography tandem mass spectrometry (HPLC-MS/MS) as previously described (39, 51). The peptide fingerprint was evaluated using ProteinPilot software 4.0 in combination with the EC958 protein database. SDS-PAGE and immunoblotting were performed as previously described (54) using a YncE-

specific antibody. Rabbit YncE antibodies were raised against purified YncE by the Walter and Eliza Hall Institute Antibody Facility as previously described (55).

Expression and purification of vaccine targets. Polyhistidine-tagged recombinant proteins were obtained by autoinduction (56) and purified by affinity and gel filtration chromatography. Briefly, BL21(DE3)(pLysS) cells harboring the LIC vectors were grown in ZYP-5052 (1 liter) overnight at 28°C at 250 rpm. Cells were harvested by centrifugation at $18,600 \times g$, resuspended in a mixture of 25 mM Tris-HCl (pH 7), 150 mM NaCl, 0.5% Triton X-100, protease inhibitor cocktail (Sigma), and DNase I (Roche), and lysed by sonication (Misonix, Inc., XL-2000; QSonica). Cell debris was removed by centrifugation, and histidine-tagged proteins were purified using Talon metal affinity resin (Clontech) preequilibrated with buffer A (25 mM HEPES [pH 7], 150 mM NaCl). Proteins bound to the resin were washed extensively with buffer B (25 mM Tris-HCl, 500 mM NaCl, 10 mM imidazole) and eluted with buffer C (25 mM Tris-HCl [pH 7], 150 mM NaCl, 250 mM imidazole). Fractions containing the eluted protein were pooled and dialyzed overnight in buffer A and further purified by gel filtration chromatography (Åkta, GE Healthcare) using a HiLoad 16/60 Superdex 75 prep-grade column (GE Healthcare) preequilibrated in buffer A.

Plasma collections and immunoassays. Blood plasma was collected from 47 urosepsis patients admitted to the Princess Alexandra Hospital (Brisbane, Australia) and 42 healthy volunteers with no recent history of UTI. Recombinant proteins (10 $\mu\text{g/ml}$) were coated onto Nunc Maxisorp flat-bottom 96-well plates (Thermo Scientific) in carbonate coating buffer (18 mM Na_2CO_3 , 450 mM NaHCO_3 [pH 9.3]) at 4°C overnight. Plates were then washed twice with phosphate-buffered saline–0.05% Tween 20 (PBST) and blocked on 5% skim milk in PBST (150 μl) for 90 min at 37°C. Plates were washed four times with PBST, and then plasma samples were added to the wells at a 1:10 dilution. Plates were incubated for 90 min at 37°C and washed four times with PBST. Peroxidase-conjugated anti-human IgG (1:30,000 dilution in 0.5% skim milk) was applied as a secondary antibody and incubated for 90 min at 37°C. Plates were washed four times with PBST before undergoing development with 3,3',5,5'-tetramethylbenzidine. Reactions were stopped with 1 M HCl. Intensity was determined using SpectraMax 190 absorbance microplate reader at 450 nm. Statistical analysis for comparisons between patient and healthy plasma was performed using an unpaired two-sample t test. A statistical significance threshold was set at $P < 0.05$.

A rabbit polyclonal antiserum was raised against purified YncE using four immunizations (400 μg recombinant protein/dose) at the Walter and Eliza Hall Institute Antibody Facility. For immunoblotting, samples were subjected to SDS-PAGE using 12% Bis-Tris gels and subsequently transferred to polyvinylidene difluoride (PVDF) microporous membrane. YncE antiserum was used as primary serum, and the secondary antibody was alkaline phosphatase-conjugated anti-rabbit IgG. Sigma Fast 5-bromo-4-chloro-3-indolylphosphate–nitroblue tetrazolium (BCIP/NBT) was used as the substrate in the detection process.

Bacteremia model of infection. A murine model of *E. coli* bacteremia, as previously described (27), was used to assess the protective efficacy of YncE as a vaccine immunogen against systemic infection. Groups of 12 C57BL/6 mice (8 to 12 weeks old) were immunized subcutaneously (s.c.) with 100 μg of YncE in 100 μl of an emulsification of PBS and complete Freund's adjuvant (Sigma) (2:1) on day 0. Booster doses of 25 μg of antigen in 100 μl of an emulsification of PBS and incomplete Freund's adjuvant (Sigma) (1:1) were administered s.c. on days 7 and 14, essentially as previously described (57). Mice were challenged intravenously (i.v.) with approximately 6.4×10^6 CFU of *E. coli* CFT073 in 200 μl of PBS via the lateral tail vein on day 21. The burden of disease was assessed at 24 h postchallenge by quantitating the bacterial loads in liver, blood, kidney, and spleen. The experiment included mock-immunized controls (which received PBS and adjuvant only) and was repeated independently.

Blood samples were also collected on days 0, 7, 14, 21, and 22 to measure YncE-specific IgG antibody titers by enzyme-linked immunosorbent assay (ELISA). Sera from immunized mice were separated from blood at $1,500 \times g$ for 10 min. Plates were coated, blocked, and washed as detailed above. Mouse serum samples were applied at a 1:2 serial dilution, starting from a 1:10 dilution. Subsequent wash, secondary antibody, and development steps were performed as previously described (54), with the exception that peroxidase-conjugated anti-mouse antibodies were applied as the secondary antibody. YncE IgG titers were defined as the logarithmic dilution that produced a significant absorbance (450 nm) in comparison to that of a blank sample.

Ethics statement. This study was performed in accordance with the ethical standards of the University of Queensland, Princess Alexandra Hospital, Griffith University, and the Helsinki Declaration. The collection of human plasma was approved, and the need for informed consent was waived, by the institutional review boards of the Princess Alexandra Hospital (research protocol 2008/264) and Griffith University (MSC/18/10/HREC). Approval for mouse infection studies was obtained from The Griffith University Animal Ethics Committee (MSC/03/15/AEC).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/mSphere.00326-16>.

Figure S1, TIF file, 2.7 MB.

Figure S2, TIF file, 0.4 MB.

Data Set S1, XLSX file, 0.1 MB.

Data Set S2, XLSX file, 0.1 MB.

Data Set S3, XLSX file, 0.1 MB.

Data Set S4, XLSX file, 0.3 MB.

Data Set S5, XLS file, 2.6 MB.

Data Set S6, XLSX file, 0.1 MB.

Data Set S7, XLSX file, 0.1 MB.

Data Set S8, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

We thank Alun Jones for expert liquid chromatography technical support. We thank all of the researchers whose collaborative spirit fosters the growth of publically available genomic resources by allowing access to their data.

This work was supported by grants from the Australian National Health and Medical Research Council (NHMRC) (GNT1012076, GNT1067455, and GNT1084889). M.A.S. is supported by an NHMRC Senior research fellowship (GNT1106930), S.A.B. by an NHMRC Career Development fellowship (GNT1090456), and G.C.U. by an Australian Research Council (ARC) Future fellowship (FT110101048). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

- Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, Paxinos EE, Sebra R, Chin CS, Iliopoulos D, Klammer A, Peluso P, Lee L, Kislyuk AO, Bullard J, Kasarskis A, Wang S, Eid J, Rank D, Redman JC, Steyert SR, Frimodt-Møller J, Struve C, Petersen AM, Krogfelt KA, Nataro JP, Schadt EE, Waldor MK. 2011. Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N Engl J Med* **365**:709–717. <http://dx.doi.org/10.1056/NEJMoa1106920>.
- Totsika M, Beatson SA, Sarkar S, Phan MD, Petty NK, Bachmann N, Szubert M, Sidjabat HE, Paterson DL, Upton M, Schembri MA. 2011. Insights into a multidrug resistant *Escherichia coli* pathogen of the globally disseminated ST131 lineage: genome analysis and virulence mechanisms. *PLoS One* **6**:e26578. <http://dx.doi.org/10.1371/journal.pone.0026578>.
- Peirano G, Bradford PA, Kazmierczak KM, Badal RE, Hackel M, Hoban DJ, Pitout JD. 2014. Global incidence of carbapenemase-producing *Escherichia coli* ST131. *Emerg Infect Dis* **20**:1928–1931. <http://dx.doi.org/10.3201/eid2011.141388>.
- Kaper JB, Nataro JP, Mobley HL. 2004. Pathogenic *Escherichia coli*. *Nat Rev Microbiol* **2**:123–140. <http://dx.doi.org/10.1038/nrmicro818>.
- Didelot X, Méric G, Falush D, Darling AE. 2012. Impact of homologous and non-homologous recombination in the genomic evolution of *Escherichia coli*. *BMC Genomics* **13**:256. <http://dx.doi.org/10.1186/1471-2164-13-256>.
- Jaureguy F, Landraud L, Passet V, Diancourt L, Frapy E, Guigon G, Carbonnelle E, Lortholary O, Clermont O, Denamur E, Picard B, Nassif X, Brisse S. 2008. Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. *BMC Genomics* **9**:560. <http://dx.doi.org/10.1186/1471-2164-9-560>.
- von Mentzer A, Connor TR, Wieler LH, Semmler T, Iguchi A, Thomson NR, Rasko DA, Joffe E, Corander J, Pickard D, Wiklund G, Svennerholm AM, Sjöling Å, Dougan G. 2014. Identification of enterotoxigenic *Escherichia coli* (ETEC) clades with long-term global distribution. *Nat Genet* **46**:1321–1326. <http://dx.doi.org/10.1038/ng.3145>.
- Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MC, Ochman H, Achtman M. 2006. Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol* **60**:1136–1151. <http://dx.doi.org/10.1111/j.1365-2958.2006.05172.x>.
- Karch H, Denamur E, Dobrindt U, Finlay BB, Hengge R, Johannes L, Ron EZ, Tønjum T, Sansonetti PJ, Vicente M. 2012. The enemy within us: lessons from the 2011 European *Escherichia coli* O104:H4 outbreak. *EMBO Mol Med* **4**:841–848. <http://dx.doi.org/10.1002/emmm.201201662>.
- Moriel DG, Rosini R, Seib KL, Serino L, Pizza M, Rappuoli R. 2012. *Escherichia coli*: great diversity around a common core. *mBio* **3**:e00118–12. <http://dx.doi.org/10.1128/mBio.00118-12>.
- Clermont O, Gordon D, Denamur E. 2015. Guide to the various phylogenetic classification schemes for *Escherichia coli* and the correspondence among schemes. *Microbiology* **161**:980–988. <http://dx.doi.org/10.1099/mic.0.000063>.
- Salipante SJ, Roach DJ, Kitzman JO, Snyder MW, Stackhouse B, Butler-Wu SM, Lee C, Cookson BT, Shendure J. 2015. Large-scale genomic sequencing of extraintestinal pathogenic *Escherichia coli* strains. *Genome Res* **25**:119–128. <http://dx.doi.org/10.1101/gr.180190.114>.
- Ahmed SA, Awosika J, Baldwin C, Bishop-Lilly KA, Biswas B, Broomall S, Chain PS, Chertkov O, Chokoshvili O, Coyne S, Davenport K, Detter JC, Dorman W, Erkkila TH, Folster JP, Frey KG, George M, Gleasner C, Henry M, Hill KK, Hubbard K, Insalaco J, Johnson S, Kitzmiller A, Krepps M, Lo CC, Luu T, McNew LA, Minogue T, Munk CA, Osborne B, Patel M, Reitenga KG, Rosenzweig CN, Shea A, Shen X, Strockbine N, Tarr C, Teshima H, van Gieson E, Verratti K, Wolcott M, Xie G, Sozhamannan S, Gibbons HS. 2012. Genomic comparison of *Escherichia coli* O104:H4 isolates from 2009 and 2011 reveals plasmid, and prophage heterogeneity, including Shiga toxin encoding phage stx2. *PLoS One* **7**:e48228. <http://dx.doi.org/10.1371/journal.pone.0048228>.
- Brzuszkiewicz E, Thürmer A, Schuldes J, Leimbach A, Liesegang H, Meyer FD, Boelter J, Petersen H, Gottschalk G, Daniel R. 2011. Genome sequence analyses of two isolates from the recent *Escherichia coli* outbreak in Germany reveal the emergence of a new pathotype: entero-aggregative-haemorrhagic *Escherichia coli* (EAHEC). *Arch Microbiol* **193**:883–891. <http://dx.doi.org/10.1007/s00203-011-0725-6>.
- Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, Prior K, Szczepanowski R, Ji Y, Zhang W, McLaughlin SF, Henkhaus JK, Leopold B, Bielaszewska M, Prager R, Brzoska PM, Moore RL, Guenther S, Rothberg JM, Karch H. 2011. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS One* **6**:e22751. <http://dx.doi.org/10.1371/journal.pone.0022751>.
- Toval F, Schiller R, Meisen I, Putze J, Kouzel IU, Zhang W, Karch H, Bielaszewska M, Mormann M, Müthing J, Dobrindt U. 2014. Characterization of urinary tract infection-associated Shiga toxin-producing *Escherichia coli*. *Infect Immun* **82**:4631–4642. <http://dx.doi.org/10.1128/IAI.01701-14>.
- Petty NK, Ben Zakour NL, Stanton-Cook M, Skippington E, Totsika M, Forde BM, Phan MD, Gomes Moriel D, Peters KM, Davies M, Rogers BA, Dougan G, Rodriguez-Baño J, Pascual A, Pitout JD, Upton M, Paterson DL, Walsh TR, Schembri MA, Beatson SA. 2014. Global dissemination of a multidrug resistant *Escherichia coli* clone. *Proc Natl Acad Sci U S A* **111**:5694–5699. <http://dx.doi.org/10.1073/pnas.1322678111>.
- Price LB, Johnson JR, Aziz M, Clabots C, Johnston B, Tchesnokova V, Nordstrom L, Billig M, Chatopadhyay S, Stegger M, Andersen PS, Pearson T, Riddell K, Rogers P, Scholes D, Kahl B, Keim P, Sokurenko EV. 2013. The epidemic of extended-spectrum-beta-lactamase-producing *Escherichia coli* ST131 is driven by a single highly pathogenic subclone, H30-Rx. *mBio* **4**:e00377-13. <http://dx.doi.org/10.1128/mBio.00377-13>.

19. Luo C, Walk ST, Gordon DM, Feldgarden M, Tiedje JM, Konstantinidis KT. 2011. Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. *Proc Natl Acad Sci U S A* **108**:7200–7205. <http://dx.doi.org/10.1073/pnas.1015622108>.
20. Walk ST, Alm EW, Gordon DM, Ram JL, Toranzos GA, Tiedje JM, Whittam TS. 2009. Cryptic lineages of the genus *Escherichia*. *Appl Environ Microbiol* **75**:6534–6544. <http://dx.doi.org/10.1128/AEM.01262-09>.
21. Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martínez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M, Latendresse M, Muñoz-Rascado L, Ong Q, Paley S, Schröder I, Shearer AG, Subhraveti P, Travers M, Weerasinghe D, Weiss V, Collado-Vides J, Gunsalus RP, Paulsen I, Karp PD. 2013. EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res* **41**:D605–D612. <http://dx.doi.org/10.1093/nar/gks1027>.
22. Phan MD, Peters KM, Sarkar S, Lukowski SW, Allsopp LP, Moriel DG, Achard ME, Totsika M, Marshall VM, Upton M, Beatson SA, Schembri MA. 2013. The serum resistome of a globally disseminated multidrug resistant uropathogenic clone. *PLoS Genet* **9**:e1003834. <http://dx.doi.org/10.1371/journal.pgen.1003834>.
23. Blattner FR, Plunkett G, III, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**:1453–1462. <http://dx.doi.org/10.1126/science.277.5331.1453>.
24. Njamkepo E, Fawal N, Tran-Dien A, Hawkey J, Strockbine N, Jenkins C, Talukder KA, Bercion R, Kuleshov K, Kolínská R, Russell JE, Kaftyreva L, Accou-Demartin M, Karas A, Vandenberg O, Mather AE, Mason CJ, Page AJ, Ramamurthy T, Bizet C, Gaman A, Carle I, Sow AG, Bouchier C, Wester AL, Lejay-Collin M, Fonkkoa MC, Hello SL, Blaser MJ, Jernberg C, Ruckly C, Mérens A, Page AL, Aslett M, Roggentin P, Fruth A, Denamur E, Venkatesan M, Bercovier H, Bodhidatta L, Chiou CS, Clermont D, Colonna B, Egorova S, Pashani GP, Ezernitchi AV, Guigon G, Harris SR, Izumiya H, Korzeniowska-Kowal A. 2016. Global phylogeography and evolutionary history of *Shigella dysenteriae* type 1. *Nat Microbiol* **1**:16027. <http://dx.doi.org/10.1038/nmicrobiol.2016.27>.
25. Ingle DJ, Tauschek M, Edwards DJ, Hocking DM, Pickard DJ, Azzopardi KI, Amarasena T, Bennett-Wood V, Pearson JS, Tamboura B, Antonio M, Ochieng JB, Oundo J, Mandomando I, Qureshi S, Ramamurthy T, Hossain A, Kotloff KL, Nataro JP, Dougan G, Levine MM, Robins-Browne RM, Holt KE. 2016. Evolution of atypical enteropathogenic *E. coli* by repeated acquisition of LEE pathogenicity island variants. *Nat Microbiol* **1**:15010. <http://dx.doi.org/10.1038/nmicrobiol.2015.10>.
26. Moriel DG, Bertoldi I, Spagnuolo A, Marchi S, Rosini R, Nesta B, Pastorello I, Corea VA, Torricelli G, Cartocci E, Savino S, Scarselli M, Dobrindt U, Hacker J, Tettelin H, Tallon LJ, Sullivan S, Wieler LH, Ewers C, Pickard D, Dougan G, Fontana MR, Rappuoli R, Pizza M, Serino L. 2010. Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* **107**:9072–9077. <http://dx.doi.org/10.1073/pnas.0915077107>.
27. Smith SN, Hagan EC, Lane MC, Mobley HLT. 2010. Dissemination and systemic colonization of uropathogenic *Escherichia coli* in a murine model of bacteremia. *mBio* **1**:e00262-10. <http://dx.doi.org/10.1128/mBio.00262-10>.
28. Zhang W, Sack DA. 2012. Progress and hurdles in the development of vaccines against enterotoxigenic *Escherichia coli* in humans. *Expert Rev Vaccines* **11**:677–694. <http://dx.doi.org/10.1586/erv.12.37>.
29. Horne C, Vallance BA, Deng W, Finlay BB. 2002. Current progress in enteropathogenic and enterohemorrhagic *Escherichia coli* vaccines. *Expert Rev Vaccines* **1**:483–493. <http://dx.doi.org/10.1586/14760584.1.4.483>.
30. García-Angulo VA, Kalita A, Torres AG. 2013. Advances in the development of enterohemorrhagic *Escherichia coli* vaccines using murine models of infection. *Vaccine* **31**:3229–3235. <http://dx.doi.org/10.1016/j.vaccine.2013.05.013>.
31. Moriel DG, Schembri MA. 2013. Vaccination approaches for the prevention of urinary tract infection. *Curr Pharm Biotechnol* **14**:967–974. <http://dx.doi.org/10.2174/1389201014666131226144824>.
32. Hacker J, Carniel E. 2001. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. *EMBO Rep* **2**:376–381. <http://dx.doi.org/10.1093/embo-reports/kve097>.
33. Kulp A, Kuehn MJ. 2010. Biological functions and biogenesis of secreted bacterial outer membrane vesicles. *Annu Rev Microbiol* **64**:163–184. <http://dx.doi.org/10.1146/annurev.micro.091208.073413>.
34. Nesta B, Valeri M, Spagnuolo A, Rosini R, Mora M, Donato P, Alteri CJ, Del Vecchio M, Buccato S, Pizzicoli A, Bertoldi I, Buzzigoli L, Tuscano G, Falduto M, Rippa V, Ashhab Y, Bensi G, Fontana MR, Seib KL, Mobley HL, Pizza M, Soriani M, Serino L. 2014. SsIE elicits functional antibodies that impair in vitro mucinase activity and in vivo colonization by both intestinal and extraintestinal *Escherichia coli* strains. *PLoS Pathog* **10**:e1004124. <http://dx.doi.org/10.1371/journal.ppat.1004124>.
35. Webb CT, Selkrig J, Perry AJ, Noinaj N, Buchanan SK, Lithgow T. 2012. Dynamic association of BAM complex modules includes surface exposure of the lipoprotein BamC. *J Mol Biol* **422**:545–555. <http://dx.doi.org/10.1016/j.jmb.2012.05.035>.
36. Baba-Dikwa A, Thompson D, Spencer NJ, Andrews SC, Watson KA. 2008. Overproduction, purification and preliminary X-ray diffraction analysis of YncE, an iron-regulated Sec-dependent periplasmic protein from *Escherichia coli*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **64**:966–969. <http://dx.doi.org/10.1107/S1744309108029515>.
37. Baars L, Ytterberg AJ, Drew D, Wagner S, Thilo C, van Wijk KJ, de Gier JW. 2006. Defining the role of the *Escherichia coli* chaperone SecB using comparative proteomics. *J Biol Chem* **281**:10024–10034. <http://dx.doi.org/10.1074/jbc.M509929200>.
38. Kagawa W, Sagawa T, Niki H, Kurumizaka H. 2011. Structural basis for the DNA-binding activity of the bacterial beta-propeller protein YncE. *Acta Crystallogr D Biol Crystallogr* **67**:1045–1053. <http://dx.doi.org/10.1107/S0907444911045033>.
39. Wurpel DJ, Moriel DG, Totsika M, Easton DM, Schembri MA. 2014. Comparative analysis of the uropathogenic *Escherichia coli* surface proteome by tandem mass-spectrometry of artificially induced outer membrane vesicles. *J Proteomics* **115**:93–106. <http://dx.doi.org/10.1016/j.jprot.2014.12.005>.
40. McHugh JP, Rodríguez-Quinoñes F, Abdul-Tehrani H, Svistunenko DA, Poole RK, Cooper CE, Andrews SC. 2003. Global iron-dependent gene regulation in *Escherichia coli*. A new mechanism for iron homeostasis. *J Biol Chem* **278**:29478–29486. <http://dx.doi.org/10.1074/jbc.M303381200>.
41. Hays MP, Ericsson AC, Yang Y, Hardwidge PR. 2016. Vaccinating with conserved *Escherichia coli* antigens does not alter the mouse intestinal microbiome. *BMC Res Notes* **9**:401. <http://dx.doi.org/10.1186/s13104-016-2208-y>.
42. Pearson WR, Wood T, Zhang Z, Miller W. 1997. Comparison of DNA sequences with protein sequences. *Genomics* **46**:24–36. <http://dx.doi.org/10.1006/geno.1997.4995>.
43. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FS. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**:1608–1615. <http://dx.doi.org/10.1093/bioinformatics/btq249>.
44. Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A. 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci* **12**:1652–1662. <http://dx.doi.org/10.1110/ps.0303703>.
45. Kelley LA, Sternberg MJ. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**:363–371. <http://dx.doi.org/10.1038/nprot.2009.2>.
46. Clermont O, Bonacorsi S, Bingen E. 2000. Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl Environ Microbiol* **66**:4555–4558. <http://dx.doi.org/10.1128/AEM.66.10.4555-4558.2000>.
47. Escobar-Páramo P, Le Menac'h A, Le Gall T, Amorin C, Gouriou S, Picard B, Skurnik D, Denamur E. 2006. Identification of forces shaping the commensal *Escherichia coli* genetic structure by comparing animal and human isolates. *Environ Microbiol* **8**:1975–1984. <http://dx.doi.org/10.1111/j.1462-2920.2006.01077.x>.
48. Clermont O, Christenson JK, Denamur E, Gordon DM. 2013. The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep* **5**:58–65. <http://dx.doi.org/10.1111/1758-2229.12019>.
49. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**:2725–2729. <http://dx.doi.org/10.1093/molbev/mst197>.
50. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomic data. *Genome Res* **19**:1131–1139. <http://dx.doi.org/10.1101/092492>.

- ative genomics. *Genome Res* **19**:1639–1645. <http://dx.doi.org/10.1101/gr.092759.109>.
51. **Moriel DG, Beatson SA, Wурpel DJ, Lipman J, Nimmo GR, Paterson DL, Schembri MA.** 2013. Identification of novel vaccine candidates against multidrug-resistant *Acinetobacter baumannii*. *PLoS One* **8**:e77631. <http://dx.doi.org/10.1371/journal.pone.0077631>.
 52. **Donnelly MI, Zhou M, Millard CS, Clancy S, Stols L, Eschenfeldt WH, Collart FR, Joachimiak A.** 2006. An expression vector tailored for large-scale, high-throughput purification of recombinant proteins. *Protein Expr Purif* **47**:446–454. <http://dx.doi.org/10.1016/j.pep.2005.12.011>.
 53. **Allsopp LP, Beloin C, Ulett GC, Valle J, Totsika M, Sherlock O, Ghigo JM, Schembri MA.** 2012. Molecular characterization of UpaB and UpaC, two new autotransporter proteins of uropathogenic *Escherichia coli* CFT073. *Infect Immun* **80**:321–332. <http://dx.doi.org/10.1128/IAI.05322-11>.
 54. **Nichols KB, Totsika M, Moriel DG, Lo AW, Yang J, Wурpel DJ, Rossiter AE, Strugnell RA, Henderson IR, Ulett GC, Beatson SA, Schembri MA.** 2016. Molecular characterisation of the vacuolating autotransporter toxin in uropathogenic *Escherichia coli*. *J Bacteriol* **198**:1487–1498. <http://dx.doi.org/10.1128/JB.00791-15>.
 55. **Kakkanat A, Totsika M, Schaale K, Duell BL, Lo AW, Phan MD, Moriel DG, Beatson SA, Sweet MJ, Ulett GC, Schembri MA.** 2015. The role of H4 flagella in *Escherichia coli* ST131 virulence. *Sci Rep* **5**:16149. <http://dx.doi.org/10.1038/srep16149>.
 56. **Studier FW.** 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**:207–234. <http://dx.doi.org/10.1016/j.pep.2005.01.016>.
 57. **Brumbaugh AR, Smith SN, Mobley HL.** 2013. Immunization with the yersiniabactin receptor, FyuA, protects against pyelonephritis in a murine model of urinary tract infection. *Infect Immun* **81**:3309–3316. <http://dx.doi.org/10.1128/IAI.00470-13>.