



(51) International Patent Classification:

A61K 39/108 (2006.01) C12N 15/31 (2006.01)
C07K 14/245 (2006.01) C12Q 1/10 (2006.01)

(21) International Application Number:

PCT/AU2017/051164

(22) International Filing Date:

24 October 2017 (24.10.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

2016904314 24 October 2016 (24.10.2016) AU

(71) Applicant: THE UNIVERSITY OF QUEENSLAND
[AU/AU]; St Lucia, Brisbane, Queensland 4072 (AU).(72) Inventors: SCHEMBRI, Mark; The University of
Queensland, St Lucia, Queensland 4072 (AU). MORIEL,
Danilo; The University of Queensland, Brisbane, Queens-
land 4072 (AU).(74) Agent: FISHER ADAMS KELLY CALLINANS; Level
6, 175 Eagle Street, Brisbane, Queensland 4000 (AU).(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ,
CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,

HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,
KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME,
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,
SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,
TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,
MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: IMMUNOGENIC PROTEIN AND METHOD OF USE

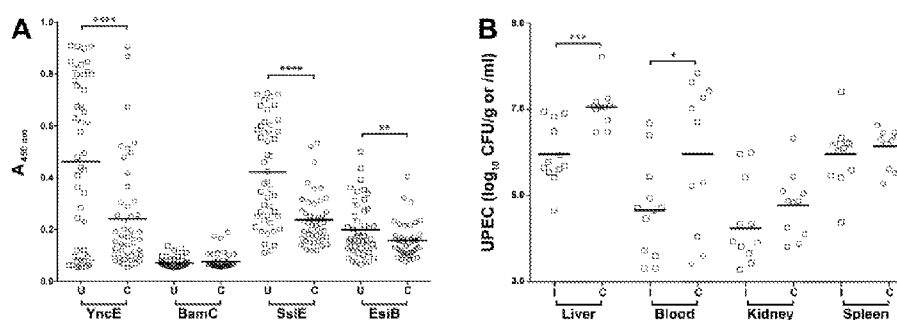


Figure 1

(57) **Abstract:** A method of eliciting an immune response in an animal includes administering an isolated protein comprising the amino acid sequence of SEQ ID NO:1; one or more immunogenic fragments, variants and/or derivatives thereof; an isolated nucleic acid encoding said isolated protein, immunogenic fragment, variant and/or derivative; and/or an antibody or antibody fragment that binds said isolated protein, immunogenic fragment, variant and/or derivative. This may immunize the animal against an *E. coli* infection and/or treat an existing *E. coli* infection.



TITLE

IMMUNOGENIC PROTEIN AND METHOD OF USE

TECHNICAL FIELD

THIS INVENTION relates to *Escherichia coli*. More particularly this invention relates
5 to an immunogenic protein that is capable of eliciting an immune response to
Escherichia coli.

BACKGROUND

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
10 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 multi-faceted pathogen, highlighted by aggressive disease outbreaks (1) and the
emergence of multidrug resistant (MDR) lineages (2, 3).

E. coli can be classified into different pathotypes according to a common set of
15 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
20 result of the independent acquisition of specific virulent 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
25 *E. coli* [EAEC] and enterohaemorrhagic *E. coli* [EHEC]) and, combined with genes
encoding resistance to antibiotics, emerged as a highly virulent lineage that infected
nearly 4000 people and caused 54 deaths (9). Since this outbreak, it has been proposed
that targeting accessory components encoded by the *E. coli* genome may be insufficient
to prevent the emergence of new pathogenic lineages, and that broader strategies
30 directed against conserved features of all strains may be more effective (10).

SUMMARY

The invention relates to an *E. coli* YncE protein suitable for eliciting an immune
response to *E. coli* in an animal. Preferably, the immune response is capable of being

directed to a plurality of different *E. coli* strains, serotypes and pathotypes. The immune response may be a protective immune response.

Accordingly, one form of the invention is broadly directed to an isolated protein comprising the amino acid sequence of SEQ ID NO:1; one or more immunogenic
5 fragments variants and/or derivatives thereof; an isolated nucleic acid encoding said isolated protein, immunogenic fragment, variant and/or derivative; and/or an antibody or antibody fragment that binds said isolated protein, immunogenic fragment, variant and/or derivative, preferably for use in eliciting an immune response in an animal.

Suitably, the immune response may be a protective immune response and/or
10 prevent or treat an *E. coli*-associated disease, disorder or condition in the animal.

In a first aspect, the invention provides a method of eliciting an immune response in an animal, said method including the step of administering to the animal: an isolated protein comprising the amino acid sequence of SEQ ID NO:1; one or more immunogenic fragments, variants and/or derivatives thereof; an isolated nucleic acid
15 encoding said isolated protein, immunogenic fragment, variant and/or derivative; and/or an antibody or antibody fragment that binds said isolated protein, immunogenic fragment, variant and/or derivative; to thereby elicit an immune response to *E. coli* in the animal.

Preferably, said immune response is a protective immune response and/or
20 confers passive immunity.

In a second aspect, the invention provides a method of immunizing an animal including the step of administering to the animal: an isolated protein comprising the amino acid sequence of SEQ ID NO:1; one or more immunogenic fragments, variants and/or derivatives thereof; an isolated nucleic acid encoding said isolated protein,
25 immunogenic fragment, variant and/or derivative; and/or an antibody or antibody fragment that binds said isolated protein, immunogenic fragment, variant and/or derivative; to thereby induce immunity to *E. coli* in the animal.

In a third aspect, the invention provides a method of treating an *E. coli*-associated disease, disorder or condition in an animal, said method including the step of
30 administering to the animal: an isolated protein comprising the amino acid sequence of SEQ ID NO:1; one or more immunogenic fragments variants and/or derivatives thereof; an isolated nucleic acid encoding said isolated protein, immunogenic fragment, variant and/or derivative; and/or an antibody or antibody fragment that binds said isolated protein, immunogenic fragment, variant and/or derivative; to thereby treat the disease,

disorder or condition in the animal.

In some embodiments of the aforementioned aspects, the isolated protein of SEQ ID NO:1, or one or more immunogenic fragments variants and/or derivatives thereof, or the isolated nucleic acid may be administered in combination with one or
5 more additional immunogens that may be specific to certain strains/pathotypes of *E. coli*, or are conserved across a number of different *E. coli* strains/pathotypes. Non-limiting examples of other such immunogens include iron-receptor proteins, fimbrial proteins, flagella, capsular antigens and O antigens.

In a fourth aspect, the invention provides a method of detecting an *E. coli*
10 bacterium, or one or more molecular components thereof, in a sample, said method including the step of detecting: an isolated protein comprising the amino acid sequence of SEQ ID NO:1; one or more immunogenic fragments, variants and/or derivatives thereof; an isolated nucleic acid encoding said isolated protein, immunogenic fragment, variant and/or derivative; and/or an antibody or antibody fragment that binds said
15 isolated protein, immunogenic fragment, variant and/or derivative; in said biological sample, which indicates the presence of *E. coli*, or one or more molecular components thereof, in the sample.

The animal according to the aforementioned aspects is preferably a mammal, or more preferably a human.

20 In a fifth aspect, the invention provides an immunogenic fragment of an isolated protein comprising the amino acid sequence of SEQ ID NO:1. Suitably, the immunogenic fragment comprises one or more T or B epitopes.

In a sixth aspect, the present invention provides an isolated nucleic acid that comprises a nucleotide sequence that encodes the immunogenic fragment of the fifth
25 aspect, or a nucleotide sequence complementary thereto.

In a seventh aspect, the invention provides a genetic construct comprising the isolated nucleic acid of the sixth aspect; operably linked or connected to one or more regulatory sequences in an expression vector.

30 In an eighth aspect, the invention provides a host cell transformed or transfected with the genetic construct of the seventh aspect.

In a ninth aspect, the invention provides a composition comprising: (a) an isolated protein comprising the amino acid sequence of SEQ ID NO:1; (b) the immunogenic fragment of the fifth aspect; (c) a variant or derivative of (a) or (b); (d) an antibody or antibody fragment that binds (a), (b) or (c); (e) an isolated nucleic acid

encoding (a), (b) or (c); (f) the genetic construct of the seventh aspect; or (g) the host cell of the eighth aspect; and a carrier, diluent or excipient.

Suitably, the composition is for use according to the method of any one of the first to fourth aspects.

5 In one embodiment, the composition for use according to the first to third aspects further comprises a carrier, diluent or excipient.

In another embodiment, the composition is suitable for detection, such as according to the fourth aspect.

Throughout this specification, unless otherwise indicated, “comprise”,
10 “comprises” and “comprising” are used inclusively rather than exclusively, so that a stated integer or group of integers may include one or more other non-stated integers or groups of integers.

By “*consist essentially of*” is meant in this context that the isolated protein or immunogenic fragment has one, two or no more than three amino acid residues in
15 addition to the recited amino acid sequence. The additional amino acid residues may occur at the N- and/or C-termini of the recited amino acid sequence, although without limitation thereto.

It will also be appreciated that the indefinite articles “a” and “an” are not to be read as singular or as otherwise excluding more than one or more than a single subject
20 to which the indefinite article refers. For example, “a” protein includes one protein, one or more proteins or a plurality of proteins.

BRIEF DESCRIPTION OF FIGURES

Figure 1. Immunoreactivity of plasma from urosepsis patients to *E. coli* vaccine antigens, and of mice to YncE. (A) Blood plasma was collected from 47
25 urosepsis patients (U) at least 4 days post admittance to hospital. IgG-specific antibodies levels were compared to 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
30 individual mice, and bars show the medians. The limit of detection was 200 CFU/g or CFU/ml. Statistically significant P values (***<0.001, *<0.05) are indicated.

Figure 2. 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 *fur* knockout. OmpA was used as an expression control.

Figure 3. Classification of the database according to the region and year of isolation, source and disease.

5 **Figure 4. A. Expression of selected targets.** TL, total cell lysates; SF, soluble fraction. **B. Murine YncE-specific IgG response following subcutaneous immunization.** Median titres with range plotted in logarithmic scale.

Figure 5. Primers used herein. The nucleotide sequences of these primers are designated herein as SEQ ID NOS:3-12, respectively.

10

BRIEF DESCRIPTION OF THE AMINO ACID AND NUCLEOTIDE SEQUENCES

SEQ ID NO:1 = amino acid sequence of *E. coli* protein YncE.

SEQ ID NO:2 = nucleotide sequence encoding SEQ ID NO:1.

15 SEQ ID NO:3 = nucleotide sequence of primer 4636.

SEQ ID NO:4 = nucleotide sequence of primer 4637.

SEQ ID NO:5 = nucleotide sequence of primer 4640.

SEQ ID NO:6 = nucleotide sequence of primer 4641.

SEQ ID NO:7 = nucleotide sequence of primer 4642.

20 SEQ ID NO:8 = nucleotide sequence of primer 4643.

SEQ ID NO:9 = nucleotide sequence of primer 4650.

SEQ ID NO:10 = nucleotide sequence of primer 4651.

SEQ ID NO:11 = nucleotide sequence of primer 5600.

SEQ ID NO:12 = nucleotide sequence of primer 5601.

25

DETAILED DESCRIPTION

The present invention is at least partly predicated on work defining core and accessory *E. coli* genes that identified YncE as a highly conserved vaccine antigen that is protective against acute *E. coli* bacteremia. The invention therefore provides methods of immunizing against, preventing and/or treating *E. coli* infections in animals by administering YncE protein to the animal. This also extends to administering immunogenic YncE fragments, variants, encoding nucleic acids and/or antibodies that bind YncE protein. A particular feature of the invention is that because YncE is “universally” expressed by a plurality of different *E. coli* strains, serotypes and

pathotypes, immunity to YncE may provide immunity to a plurality of different *E. coli* strains, serotypes and pathotypes.

Accordingly, a broad form of the invention is broadly directed to an isolated protein comprising the amino acid sequence of SEQ ID NO:1; one or more immunogenic fragments variants and/or derivatives thereof; an isolated nucleic acid encoding said isolated protein, immunogenic fragment, variant and/or derivative; and/or an antibody or antibody fragment that binds said isolated protein, immunogenic fragment, variant and/or derivative, for use in eliciting an immune response in an animal.

For the purposes of this invention, by “*isolated*” is meant material that has been removed from its natural state or otherwise been subjected to human manipulation. Isolated material may be substantially or essentially free from components that normally accompany it in its natural state, or may be manipulated so as to be in an artificial state together with components that normally accompany it in its natural state. Isolated material may be enriched, partially purified or in a substantially pure form. Isolated material may be in native, chemical synthetic or recombinant form.

By “*protein*” is meant an amino acid polymer. The amino acids may be natural or non-natural amino acids, D- or L-amino acids as are well understood in the art.

The term “*protein*” includes and encompasses “*peptide*”, which is typically used to describe a protein having no more than fifty (50) amino acids and “*polypeptide*”, which is typically used to describe a protein having more than fifty (50) amino acids.

A non-limiting example of an *E. coli* YncE protein comprises the amino acid sequence of SEQ ID NO:1:

```
mhlrlhlfssr lrgsl1llgsl lvvssfsta aeemlrkavg kgayemaysq qenalwlats
qsrklldkgyv vyrlpvtle vtqaihndlk pfgatinntt qtlwfgntvn savtaidakt
gevkgrlvld drkrteevrp lqprelvadd atntvyisgi gkesviwvvd ggniklktai
qntgkmstgl aldsegkrly ttnadgelit idtadnkils rkkllddgke hffinisldt
arqgrafitds kaaevlvvdt rngnilakva apeslavlfm parneayvth rqagkvsvld
aksykvvktf dtpthpnsa lsadgktylv svkqkstkqq eatqpddvir ial
```

(SEQ ID NO:1)

A “*fragment*” is a segment, domain, portion or region of a protein, which constitutes less than 100% of the amino acid sequence of the protein.

In general, fragments may comprise, consist essentially of, or consist of, up to 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 200, 250, 280, 300 or 320 amino acids of an amino acid sequence. Suitably, the fragment is an immunogenic fragment.

In the context of the present invention, the term “*immunogenic*” as used herein indicates the ability or potential of a protein, fragment or variant to elicit or generate an immune response to one or a plurality of strains of *E. coli*, or molecular components thereof, upon administration of the protein to an animal. It is envisaged that the immune response may be either B-lymphocyte or T-lymphocyte mediated, or a combination thereof. Advantageously, by “*immunogenic*” is meant capable of eliciting a B-lymphocyte response, although is not limited thereto, preferably including an antibody response, such as an IgG response.

Accordingly, in a preferred form a fragment may comprise at least one T cell epitope and/or at least one B cell epitope.

As used herein, a protein “*variant*” shares a definable amino acid sequence relationship with an isolated protein or immunogenic fragment disclosed herein. Preferably, protein variants share at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity with the amino acid sequence set forth in SEQ ID NO:1, or with a fragment of SEQ ID NO:1 as hereinbefore described. Suitably, the variant is immunogenic, as hereinbefore defined.

The “*variant*” proteins or fragments disclosed herein have one or more amino acids deleted, added or substituted by different amino acids. It is well understood in the art that some amino acids may be substituted or deleted without changing the activity of the immunogenic fragment and/or protein (conservative substitutions).

The term “*variant*” also includes isolated proteins or fragments thereof disclosed herein, produced from, or comprising amino acid sequences of, naturally occurring allelic variants and orthologs (*e.g.* from various strains, pathotypes or serotypes of *E. coli*) and synthetic variants, such as produced *in vitro* using mutagenesis techniques.

Suitably, protein variants are immunogenic, as hereinbefore described.

Terms used generally herein to describe sequence relationships between respective proteins and nucleic acids include “*comparison window*”, “*sequence identity*”, “*percentage of sequence identity*” and “*substantial identity*”. Because respective nucleic acids/proteins may each comprise (1) only one or more portions of a complete nucleic acid/protein sequence that are shared by the nucleic acids/proteins, and (2) one or more portions which are divergent between the nucleic acids/proteins, sequence comparisons are typically performed by comparing sequences over a “comparison window” to identify and compare local regions of sequence similarity. A

“*comparison window*” refers to a conceptual segment of typically 6, 9 or 12 contiguous residues that is compared to a reference sequence. The comparison window may comprise additions or deletions (*i.e.*, gaps) of about 20% or less as compared to the reference sequence for optimal alignment of the respective sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by computerised implementations of algorithms (Geneworks program by Intelligenetics; GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Drive Madison, WI, USA, incorporated herein by reference) or by inspection and the best alignment (*i.e.* resulting in the highest percentage homology over the comparison window) generated by any of the various methods selected. Reference also may be made to the BLAST family of programs as for example disclosed by Altschul *et al.*, 1997, Nucl. Acids Res. **25** 3389, which is incorporated herein by reference. A detailed discussion of sequence analysis can be found in Unit 19.3 of CURRENT PROTOCOLS IN MOLECULAR BIOLOGY Eds. Ausubel *et al.* (John Wiley & Sons Inc NY, 1995-1999).

The term “*sequence identity*” is used herein in its broadest sense to include the number of exact nucleotide or amino acid matches having regard to an appropriate alignment using a standard algorithm, having regard to the extent that sequences are identical over a window of comparison. Thus, a “*percentage of sequence identity*” is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. For example, “*sequence identity*” may be understood to mean the “match percentage” calculated by the DNASIS computer program (Version 2.5 for windows; available from Hitachi Software engineering Co., Ltd., South San Francisco, California, USA).

Derivatives of SEQ ID NO:1 and fragments and variants are also provided. Suitably, said derivatives are immunogenic, as hereinbefore described.

As used herein, “*derivative*” proteins have been altered, for example by conjugation or complexing with other chemical moieties, by post-translational modification (*e.g.* phosphorylation, acetylation and the like), modification of glycosylation (*e.g.* adding, removing or altering glycosylation) and/or inclusion of

additional amino acid sequences as would be understood in the art.

Additional amino acid sequences may include fusion partner amino acid sequences which create a fusion protein. By way of example, fusion partner amino acid sequences may assist in detection and/or purification of the isolated fusion protein.

5 Non-limiting examples include metal-binding (*e.g.* polyhistidine) fusion partners, maltose binding protein (MBP), Protein A, glutathione S-transferase (GST), fluorescent protein sequences (*e.g.* GFP), epitope tags such as myc, FLAG and haemagglutinin tags.

Other derivatives contemplated by the invention include, but are not limited to, modification to amino acid side chains, incorporation of non-natural amino acids and/or
10 their derivatives during peptide, or protein synthesis and the use of crosslinkers and other methods which impose conformational constraints on the immunogenic proteins, fragments and variants of the invention.

In this regard, the skilled person is referred to Chapter 15 of CURRENT PROTOCOLS IN PROTEIN SCIENCE, Eds. Coligan *et al.* (John Wiley & Sons NY
15 1995-2008) for more extensive methodology relating to chemical modification of proteins.

The isolated immunogenic proteins, fragments and/or derivatives of the present invention may be produced by any means known in the art, including but not limited to, chemical synthesis, recombinant DNA technology and proteolytic cleavage to produce
20 peptide fragments.

Chemical synthesis is inclusive of solid phase and solution phase synthesis. Such methods are well known in the art, although reference is made to examples of chemical synthesis techniques as provided in Chapter 9 of SYNTHETIC VACCINES Ed. Nicholson (Blackwell Scientific Publications) and Chapter 15 of CURRENT
25 PROTOCOLS IN PROTEIN SCIENCE Eds. Coligan *et al.*, (John Wiley & Sons, Inc. NY USA 1995-2008). In this regard, reference is also made to International Publication WO 99/02550 and International Publication WO 97/45444.

Recombinant proteins and immunogenic fragments may be conveniently prepared by a person skilled in the art using standard protocols as for example described
30 in Sambrook *et al.*, MOLECULAR CLONING. A Laboratory Manual (Cold Spring Harbor Press, 1989), in particular Sections 16 and 17; CURRENT PROTOCOLS IN MOLECULAR BIOLOGY Eds. Ausubel *et al.*, (John Wiley & Sons, Inc. NY USA 1995-2008), in particular Chapters 10 and 16; and CURRENT PROTOCOLS IN PROTEIN SCIENCE Eds. Coligan *et al.*, (John Wiley & Sons, Inc. NY USA 1995-

2008), in particular Chapters 1, 5 and 6.

Alternatively, fragments can be produced by digestion of a polypeptide with proteinases such as endoLys-C, endoArg-C, endoGlu-C and V8-protease. The digested fragments can be purified by chromatographic techniques as are well known in the art.

5 As used herein, an “antibody” may be, or comprise, an antigen-binding protein encoded by the immunoglobulin gene complex inclusive of IgA, IgD, IgE and IgG antibodies. Antibodies may be polyclonal or monoclonal, native or recombinant. Well-known protocols applicable to antibody production, purification and use may be found, for example, in Chapter 2 of Coligan *et al.*, CURRENT PROTOCOLS IN
10 IMMUNOLOGY (John Wiley & Sons NY, 1991-1994) and Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual*, Cold Spring Harbor, Cold Spring Harbor Laboratory, 1988, which are both herein incorporated by reference.

Generally, antibodies of the invention bind to or conjugate with an isolated protein, fragment, variant, or derivative of the invention. For example, the antibodies
15 may be polyclonal antibodies. Such antibodies may be prepared for example by administering an *E. coli* YncE protein, fragment, variant or derivative to a production species, such as mice or rabbits, to obtain polyclonal antisera. Methods of producing polyclonal antibodies are well known to those skilled in the art. Exemplary protocols which may be used are described for example in Coligan *et al.*, CURRENT
20 PROTOCOLS IN IMMUNOLOGY, *supra*, and in Harlow & Lane, 1988, *supra*.

Monoclonal antibodies may be produced using the standard method as for example, described in an article by Köhler & Milstein, 1975, Nature **256**, 495, which is herein incorporated by reference, or by more recent modifications thereof as for example, described in Coligan *et al.*, CURRENT PROTOCOLS IN IMMUNOLOGY,
25 *supra* by immortalizing spleen or other antibody producing cells derived from a production species which has been inoculated with one or more of the isolated proteins, fragments, variants or derivatives of the invention.

The invention also includes within its scope antibody fragments, such as Fc, Fab or F(ab)₂ fragments of the polyclonal or monoclonal antibodies referred to above.
30 Alternatively, the antibodies may comprise single chain Fv antibodies (scFvs) against the peptides of the invention. Such scFvs may be prepared, for example, in accordance with the methods described respectively in United States Patent No 5,091,513, European Patent No 239,400 or the article by Winter & Milstein, 1991, Nature 349:293. Antibodies may also include multivalent recombinant antibody fragments, such as

diabodies, triabodies and/or tetrabodies, comprising a plurality of scFvs, as well as dimerisation-activated demibodies (*e.g.*, WO/2007/062466). By way of example, such antibodies may be prepared in accordance with the methods described in Holliger et al., 1993 Proc Natl Acad Sci USA 90:6444-6448; or in Kipriyanov, 2009 Methods Mol Biol
5 562:177-93 and herein incorporated by reference in their entirety.

For aspects of the invention relating to detection of *E. coli* and/or YncE proteins such as in biological samples, the antibody or antibody fragment may be labelled. By way of example only, the antibody or antibody fragment may be labeled with biotin, an enzyme, fluorophore, radionuclide or other label. Alternatively, the antibody or
10 antibody fragment is unlabeled and a secondary antibody comprises a label. By way of example, the enzyme may be horseradish peroxidase (HRP), alkaline phosphatase (AP), β -galactosidase or glucose oxidase, although without limitation thereto. In embodiments where the antibody or antibody fragment or the secondary antibody is labeled with an enzyme, an appropriate substrate may include
15 diaminobenzidine (DAB), permanent red, 3-ethylbenzthiazoline sulfonic acid (ABTS), 5-bromo-4-chloro-3-indolyl phosphate (BCIP), nitro blue tetrazolium (NBT), 3,3',5,5'-tetramethyl benzidine (TNB) and 4-chloro-1-naphthol (4-CN), although without limitation thereto. A non-limiting example of a chemiluminescent substrate is LuminolTM, which is oxidized in the presence of HRP and hydrogen peroxide to form an
20 excited state product (3-aminophthalate). By way of example, the fluorophore may be, for example, fluorescein isothiocyanate (FITC), Alexa dyes, tetramethylrhodamine isothiocyanate (TRITL), allophycocyanin (APC), Texas Red, Cy5, Cy3, or R-Phycoerythrin (RPE) as are well known in the art.

An aspect of the invention provides a method of eliciting an immune response in
25 an animal, said method including the step of administering to the animal: an isolated protein comprising the amino acid sequence of SEQ ID NO:1; one or more immunogenic fragments, variants and/or derivatives thereof; an isolated nucleic acid encoding said isolated protein, immunogenic fragment, variant and/or derivative; and/or an antibody or antibody fragment that binds said isolated protein, immunogenic
30 fragment, variant and/or derivative; to thereby elicit an immune response to *E. coli* in the animal.

Another aspect of the invention provides a method of immunizing an animal including the step of administering to the animal: an isolated protein comprising the amino acid sequence of SEQ ID NO:1; one or more immunogenic fragments, variants

and/or derivatives thereof; an isolated nucleic acid encoding said isolated protein, immunogenic fragment, variant and/or derivative; and/or an antibody or antibody fragment that binds said isolated protein, immunogenic fragment, variant and/or derivative; to thereby induce immunity to *E. coli* in the animal.

- 5 As used herein, “elicits an immune response” indicates the ability or potential of a protein, fragment or variant, antibody or encoding nucleic to elicit or generate an immune response to one or a plurality of strains of *E. coli*, or molecular components thereof, upon administration of to an animal. Suitably, the immune response may be either B-lymphocyte or T-lymphocyte mediated, or a combination thereof.
- 10 Advantageously, the immune response includes a B-lymphocyte response, although is not limited thereto, preferably including an antibody response, such as an IgG response.

- As used herein, “immunize” and “immunization” refer to administering a protein, immunogenic fragment, variant, derivative, antibody or encoding nucleic acid to elicit or potentiate a protective immune response to *E. coli*. By “protective immune
- 15 response” is meant an immune response that is sufficient to prevent or at least reduce the severity or symptoms of an *E. coli* infection in an animal. It will be appreciated that this definition includes “passive immunization” typically elicited or conferred by the administration of an antibody or antibody fragment to an animal. As will be described hereinafter in the Examples, the presence of an IgG response by the protein of SEQ ID
- 20 NO:1 was an indicator that a protective immune response had been elicited by the protein. It is therefore proposed that administration of antibodies (*e.g.* IgG) that bind the protein of SEQ ID NO:1 may confer passive immunity.

- A third aspect of the invention provides a method of treating an *E. coli*-associated disease, disorder or condition in an animal, said method including the step of
- 25 administering to the animal: an isolated protein comprising the amino acid sequence of SEQ ID NO:1; one or more immunogenic fragments variants and/or derivatives thereof; an isolated nucleic acid encoding said isolated protein, immunogenic fragment, variant and/or derivative; and/or an antibody or antibody fragment that binds said isolated protein, immunogenic fragment, variant and/or derivative; to thereby treat the disease,
- 30 disorder or condition in the animal.

As used herein, “treating” (or “treat” or “treatment”) refers to a therapeutic intervention that ameliorates a sign or symptom of an *E. coli*-associated disease, disorder or condition after it has begun to develop. The term “ameliorating,” with reference to an *E. coli*-associated disease, disorder or condition, refers to any observable

beneficial effect of the treatment. Treatment need not be absolute to be beneficial to the subject. The beneficial effect can be determined using any methods or standards known to the ordinarily skilled artisan.

As used herein, “*preventing*” (or “*prevent*” or “*prevention*”) refers to a course of
5 action (such as administering a composition comprising a therapeutically effective amount of one or more immunogenic proteins and/or a fragment, variant or derivative thereof of the present invention) initiated prior to the onset of a symptom, aspect, or characteristic of an *E. coli*-associated disease, disorder or condition, so as to prevent or reduce the symptom, aspect, or characteristic. It is to be understood that such
10 preventing need not be absolute to be beneficial to a subject. A “*prophylactic*” treatment is a treatment administered to a subject who does not exhibit signs of an *E. coli*-associated disease, disorder or condition, or exhibits only early signs for the purpose of decreasing the risk of developing a symptom or clinical characteristic or outcome of the *E. coli*-associated disease, disorder or condition.

15 In the context of the present invention, by “*E. coli*-associated disease, disorder or condition” is meant any clinical pathology, illness, trauma or adverse health consequence resulting from exposure to, or infection by, any strain, pathotype or serotype of *E. coli*. Non-limiting examples of *E. coli*-associated diseases, disorders or conditions include extra-intestinal infections such as urinary tract infection, sepsis and
20 meningitis; and intestinal infections that lead to watery/bloody diarrhea, hemorrhagic colitis and hemolytic uremic syndrome.

Pathogenic *E. coli* can be classified into two (2) major subgroups: extra-intestinal pathogenic *E. coli* (ExPEC) and intestinal pathogenic *E. coli* (InPEC). ExPEC strains are responsible for diseases, disorders and conditions such as urinary tract
25 infections (UTIs), sepsis and meningitis and are classified as uropathogenic *E. coli* (UPEC) or neonatal meningitis-associated *E. coli* (NMEC). There are many pathotypes within the InPEC group (e.g. enterotoxigenic *E. coli* (ETEC) and enteroaggregative *E. coli* (EAEC)) which cause diarrheagenic infections and infections of the intestinal tract.

It will therefore be appreciated that in some embodiments the immunization
30 and/or treatment methods disclosed herein may be efficacious against a plurality of different *E. coli* strains, pathotypes or serotypes such as described above. In this context, the plurality of different *E. coli* strains, pathotypes or serotypes may include 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50, 100, 200, 300, 400, 450, 500, 1000, 5,000, 10,000 or more different *E. coli* strains, pathotypes or serotypes. A full listing of *E. coli* strains

that express YncE may be determined with reference to Tables 2-5.

In another aspect, the invention provides a composition comprising: (a) an isolated protein comprising the amino acid sequence of SEQ ID NO:1; (b) the immunogenic fragment of the fifth aspect; (c) a variant or derivative of (a) or (b); (d) an antibody or antibody fragment that binds (a), (b) or (c); (e) an isolated nucleic acid encoding (a), (b) or (c); (f) the genetic construct of the seventh aspect; or (g) the host cell of the eighth aspect..

In some embodiments, the composition may be a “vaccine” for eliciting a protective immune response to *E. coli*.

Typically, the composition comprises a carrier, diluent or excipient.

By “*carrier, diluent or excipient*” is generally meant a solid or liquid filler, diluent, solvent, vehicle or encapsulating substance that may be safely used in systemic administration. Depending upon the particular route of administration, a variety of carriers, well known in the art may be used. These carriers may be selected from a group including sugars, starches, cellulose and its derivatives, malt, gelatine, talc, calcium sulfate, vegetable oils, synthetic oils, polyols, alginic acid, phosphate buffered solutions, emulsifiers, isotonic saline and salts such as mineral acid salts including hydrochlorides, bromides and sulfates, organic acids such as acetates, propionates and malonates and pyrogen-free water.

A useful general reference describing carriers, diluents and excipients is Remington’s Pharmaceutical Sciences (Mack Publishing Co. N.J. USA, 1991) which is incorporated herein by reference.

In particular embodiments, the carrier diluent or excipient may include carriers, diluents and/or excipients that have immunological activity, or facilitate immunological activity. For example, these may include: thyroglobulin; albumins such as human serum albumin; toxins, toxoids or any mutant crossreactive material (CRM) of the toxin from tetanus, diphtheria, pertussis, *Pseudomonas*, *E. coli*, *Staphylococcus*, and *Streptococcus*; polyamino acids such as poly(lysine:glutamic acid); influenza; Rotavirus VP6, Parvovirus VP1 and VP2; hepatitis B virus core protein; hepatitis B virus recombinant vaccine and the like. Alternatively, a fragment or epitope of a carrier protein or other immunogenic protein may be used. For example, a T cell epitope of a bacterial toxin, toxoid or CRM may be used. In this regard, reference may be made to U.S. Patent No 5,785,973 which is incorporated herein by reference.

The composition may further comprise an adjuvant as is well known in the art.

As will be understood in the art, an "*adjuvant*" is, or comprises, one or more substances that enhance the immunogenicity and efficacy of a composition, such as a vaccine. Non-limiting examples of suitable adjuvants include squalane and squalene (or other oils of plant or animal origin); block copolymers; detergents such as Tween®-80; Quil®

5 A, mineral oils such as Drakeol or Marcol, vegetable oils such as peanut oil; *Corynebacterium*-derived adjuvants such as *Corynebacterium parvum*; *Propionibacterium*-derived adjuvants such as *Propionibacterium acne*; *Mycobacterium bovis* (Bacille Calmette and Guérin or BCG); *Bordetella pertussis* antigens; tetanus toxoid; diphtheria toxoid; surface active substances such as hexadecylamine,

10 octadecylamine, octadecyl amino acid esters, lysolecithin, dimethyldioctadecylammonium bromide, *N,N*-dioctadecyl-*N'*, *N'*bis(2-hydroxyethyl-propanediamine), methoxyhexadecylglycerol, and pluronic polyols; polyamines such as pyran, dextran sulfate, poly IC carbopol; peptides such as muramyl dipeptide and derivatives, dimethylglycine, tuftsin; oil emulsions; and mineral gels such as aluminum

15 phosphate, aluminum hydroxide or alum; interleukins such as interleukin 2 and interleukin 12; monokines such as interleukin 1; tumour necrosis factor; interferons such as gamma interferon; combinations such as saponin-aluminium hydroxide or Quil-A aluminium hydroxide; liposomes; ISCOM® and ISCOMATRIX® adjuvant; mycobacterial cell wall extract; synthetic glycopeptides such as muramyl dipeptides or

20 other derivatives; Avridine; Lipid A derivatives; dextran sulfate; DEAE-Dextran alone or with aluminium phosphate; carboxypolymethylene such as Carbopol' EMA; acrylic copolymer emulsions such as Neocryl A640 (e.g. U.S. Pat. No. 5,047,238); water in oil emulsifiers such as Montanide ISA 720; poliovirus, vaccinia or animal poxvirus proteins; or mixtures thereof and immunostimulatory DNA such as CpG

25 oligonucleotides and Toll receptor agonists.

Any suitable procedure is contemplated for producing compositions, such as vaccine compositions. Exemplary procedures include, for example, those described in New Generation Vaccines (1997, Levine *et al.*, Marcel Dekker, Inc. New York, Basel, Hong Kong), which is incorporated herein by reference.

30 Any safe route of administration may be employed for providing an animal with the composition of the invention. For example, oral, rectal, parenteral, sublingual, buccal, intravenous, intranasal, intra-articular, intra-muscular, intra-dermal, subcutaneous, inhalational, intraocular, intraperitoneal, intracerebroventricular and transdermal administration may be employed.

Dosage forms include tablets, dispersions, suspensions, injections, solutions, syrups, troches, capsules, nasal sprays, suppositories, aerosols, transdermal patches and the like. These dosage forms may also include injecting or implanting controlled releasing devices designed specifically for this purpose or other forms of implants
5 modified to act additionally in this fashion. Controlled release of the therapeutic agent may be effected by coating the same, for example, with hydrophobic polymers including acrylic resins, waxes, higher aliphatic alcohols, polylactic and polyglycolic acids and certain cellulose derivatives such as hydroxypropylmethyl cellulose. In addition, the controlled release may be effected by using other polymer matrices,
10 liposomes and/or microspheres.

Compositions of suitable for administration may be presented as discrete units such as capsules, caplets, sachets, functional foods/feeds or tablets, or as a powder or granules or as a solution or a suspension in an aqueous liquid, a non-aqueous liquid, an oil-in-water emulsion or a water-in-oil liquid emulsion.

15 The composition may be administered in a manner compatible with the dosage formulation, and in such amount as is immunologically effective. The dose administered to an animal should be sufficient to effect a beneficial response in an animal over an appropriate period of time. The quantity of agent(s) to be administered may depend on the animal to be treated inclusive of the age, sex, weight and general
20 health condition thereof, factors that will depend on the judgement of the practitioner.

It will also be appreciated that the methods of immunization, prevention and/or treatment of an *E. coli*-associated disease, disorder or condition, and compositions therefor, may include administration of an isolated nucleic acid encoding an isolated protein having the amino acid sequence of SEQ ID NO:1, or a fragment, variant or
25 derivative thereof.

The term “*nucleic acid*” as used herein designates single- or double-stranded DNA and RNA. DNA includes genomic DNA and cDNA. RNA includes mRNA, RNA, RNAi, siRNA, cRNA and autocatalytic RNA. Nucleic acids may also be DNA-RNA hybrids. A nucleic acid comprises a nucleotide sequence which typically includes
30 nucleotides that comprise an A, G, C, T or U base. However, nucleotide sequences may include other bases such as inosine, methylcytosine, methylinosine, methyladenosine and/or thiouridine, although without limitation thereto.

An embodiment of an isolated nucleic acid encoding an isolated protein having the amino acid sequence of SEQ ID NO:1 comprises the nucleotide sequence of SEQ

ID NO:2.

atgcatttac gtcactgtt ttcatcgcgc ctgcgtgggt cattactgtt aggttcattg
 cttgttgctt catcattcag tacgcaggcc gcagaagaaa tgctgcgtaa agcggtaggt
 aaaggtgcct acgaaatggc ttatagccag caagaaaacg cgctgtggct tgccacttcg
 5 caaagccgca aactggataa aggcggcgtg gttatcgtc ttgatccggt tactctggaa
 gtgacgcagg cgatccataa cgatctcaag ccgtttgggt ccaccatcaa taacacgact
 cagacgttgt ggtttggtaa caccgtaa atgtgcgtca cggcgataga tgccaaaacg
 ggcaagtga aaggtcgtct ggtgctggat gatcgtaagc gcacggaaga ggtacgcccg
 ttgaaccac gtgagctggt agctgatgat gccacgaaca ccgtttacat cagtggattt
 10 ggtaaagaga gcgtgatttg ggtcgtgat ggcgagaata tcaaactgaa aaccgccatc
 cagaacaccg gtaaatgag taccggctctg gcgctggata gcaaaggcaa acgtctttac
 accactaacg ctgacggcga attgattacc atcgacaccg ccgacaataa aatctcagc
 cgtaaaaagc tgctggatga cggcaaagag cactcttta tcaacatcag ccttgatacc
 accaggcagc gtgcatttat caccgattct aaagcggcag aagtgttagt ggtcgatacc
 15 cgtaatggca atattctggc gaaggttgcg gcaccggaat cactggctgt gctgtttaac
 ccagcgcgta atgaagccta cgtgacgcat cgtcaggcag gtaaagtcag tgtgattgac
 gcgaaaagct ataaagtgtt gaaaacgttc gatacgccga ctcatccgaa cagcctggcg
 ctgtctgccg atggcaaac gctgtatgtc agtgtgaaac aaaaatccac taaacagcag
 gaagctaccc agccggacga tgtgattcgt attgcgctgtaa
 20 (SEQ ID NO:2)

It will also be understood that nucleic acids may be in the form of nucleic acid
 variants or fragments of SEQ ID NO:2. In some embodiments, nucleic acid variants and
 fragments of SEQ ID NO:2 respectively encode protein variants and fragments of SEQ
 ID NO:1. In other embodiments, nucleic acid variants have at least 70%, 75%, 80%,
 25 85%, 90% 95%, 96%, 97%, 98% or 99% nucleotide sequence identity to SEQ ID
 NO:2. In yet other embodiments, nucleic acid variants hybridize under high stringency
 conditions with the nucleotide sequence of SEQ ID NO:2. Non-limiting examples of
 high stringency hybridization and wash conditions may be found in Sambrook *et al.*,
 MOLECULAR CLONING. A Laboratory Manual (Cold Spring Harbor Press, 1989),
 30 *supra* and Chapter 2 of CURRENT PROTOCOLS IN MOLECULAR BIOLOGY Eds.
 Ausubel *et al.*, *supra*.

In further embodiments, nucleic acid variants are allelic variants, orthologues or
 other polymorphic variants of the *E. coli* YncE nucleotide sequence of SEQ ID NO:2.

In some embodiments, nucleic acid fragments may comprise at least 20, 30, 40,

50, 100, 150, 100, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950 and up to 1000 contiguous nucleotides of SEQ ID NO:2.

For the purposes of delivery of a nucleic acid, it is preferable that the nucleic acid is in a genetic construct that comprises the isolated nucleic acid operably linked or
5 connected to one or more other genetic components.

Broadly, the genetic construct is in the form of, or comprises genetic components of, a plasmid, bacteriophage, a cosmid, a yeast or bacterial artificial chromosome as are well understood in the art. Genetic constructs may be suitable for maintenance and propagation of the isolated nucleic acid in bacteria or other host cells,
10 for manipulation by recombinant DNA technology and/or expression of the nucleic acid or an encoded protein of the invention.

For the purposes of host cell expression, the genetic construct is an expression construct. Suitably, the expression construct comprises the nucleic acid of the invention operably linked to one or more additional sequences in an expression vector. An
15 “*expression vector*” may be either a self-replicating extra-chromosomal vector such as a plasmid, or a vector that integrates into a host genome.

By “*operably linked*” is meant that said additional nucleotide sequence(s) is/are positioned relative to the nucleic acid of the invention preferably to initiate, regulate or otherwise control transcription.

20 Regulatory nucleotide sequences will generally be appropriate for the host cell used for expression. Numerous types of appropriate expression vectors and suitable regulatory sequences are known in the art for a variety of host cells.

Typically, said one or more regulatory nucleotide sequences may include, but are not limited to, promoter sequences, leader or signal sequences, ribosomal binding
25 sites, transcriptional start and termination sequences, translational start and termination sequences, and enhancer or activator sequences.

Constitutive or inducible promoters as known in the art are contemplated by the invention.

The expression construct may also include an additional nucleotide sequence
30 encoding a fusion partner (typically provided by the expression vector) so that the recombinant protein is expressed as a fusion protein, as hereinbefore described.

The expression construct may also include an additional nucleotide sequence encoding a selection marker such as amp^R , neo^R or kan^R , although without limitation thereto.

In particular embodiments relating to delivery of isolated nucleic acids (*e.g.* “DNA vaccination”), the expression construct may be in the form of plasmid DNA, suitably comprising a promoter operable in an animal cell (*e.g.* a CMV promoter). A review that discusses DNA plasmid vaccination is Williams, 2013, *Vaccine* **1** 225.

5 Reference is also made to DNA Vaccines, Methods and Protocols, Second Edition (Volume 127 of Methods in Molecular Medicine series, Humana Press, 2006).

In other embodiments, the nucleic acid may be in the form of a viral construct such as an adenoviral, vaccinia, lentiviral or adeno-associated viral vector. Non-limiting examples of viral constructs for immunization are provided in Hu *et al.*, 2011, *Immunol.*

10 Rev. **239** 45 and Nieto & Salvetti, 2014, *Front. Immunol.* **5** 5.

In a further aspect, the invention provides a host cell transformed with a nucleic acid molecule or a genetic construct described herein.

Suitable host cells for expression may be prokaryotic or eukaryotic. For example, suitable host cells may include but are not limited to mammalian cells (*e.g.* *HeLa*, *HEK293T*, *Jurkat cells*), yeast cells (*e.g.* *Saccharomyces cerevisiae*), insect cells (*e.g.* *Sf9*, *Trichoplusia ni*) utilized with or without a baculovirus expression system, plant cells (*e.g.* *Chlamydomonas reinhardtii*, *Phaeodactylum tricornutum*) or bacterial cells, such as *E. coli*. Introduction of genetic constructs into host cells (whether prokaryotic or eukaryotic) is well known in the art, as for example described in

15
20 CURRENT PROTOCOLS IN MOLECULAR BIOLOGY Eds. Ausubel *et al.*, (John Wiley & Sons, Inc. 1995-2009), in particular Chapters 9 and 16.

It will also be understood that the isolated protein of SEQ ID NO:1, or one or more immunogenic fragments, variants and/or derivatives thereof, or the isolated nucleic acid may be administered in combination with one or more additional

25 immunogens that may be specific to certain strains/pathotypes of *E. coli*, or are conserved across a number of *E. coli* strains/pathotypes. Non-limiting examples of other such immunogens include iron-receptor proteins, fimbrial proteins, flagella, capsular antigens and O antigens.

A further aspect of the invention provides a method of detecting an *E. coli*

30 bacterium, or one or more molecular components thereof, in a sample, said method including the step of detecting: an isolated protein comprising the amino acid sequence of SEQ ID NO:1; one or more immunogenic fragments, variants and/or derivatives thereof; an isolated nucleic acid encoding said isolated protein, immunogenic fragment, variant and/or derivative; and/or an antibody or antibody fragment that binds said

isolated protein, immunogenic fragment, variant and/or derivative; in said biological sample, which indicates the presence of *E. coli*, or one or more molecular components thereof, in the sample.

5 The sample may be obtained from an animal, such as a fluid or cellular sample inclusive of blood, serum, plasma, urine, feces, biopsy, swab or smear, although without limitation thereto. The presence of *E. coli*, or one or more molecular components thereof, in the sample may indicate that the animal is, or has been, infected with *E. coli*.

10 In other embodiments, the sample may be an environmental sample such as water or effluent or a food or beverage sample. The presence of *E. coli*, or one or more molecular components thereof, in the sample may indicate that the environmental, food or beverage sample, or has been, contaminated with *E. coli*.

The method may use protein- and/or nucleic acid-based detection methods as are well known in the art. Protein-based detection preferably includes using antibodies or antibody fragments that bind SEQ ID NO:1, one or more immunogenic fragments, 15 variants and/or derivatives thereof. Alternatively, protein-based detection may identify antibodies (*e.g.* serum IgG) to SEQ ID NO:1, one or more immunogenic fragments, variants and/or derivatives thereof in a sample obtained from an animal, which indicates that the animal is, or has been, infected with *E. coli*. Non-limiting examples of antibody-based detection includes ELISA, radioimmunoassay, immunoblotting, and 20 immunoprecipitation, although without limitation thereto.

By way of example, nucleic acid detection may be performed using one or more of nucleic acid sequence amplification (*e.g.* polymerase chain reaction, rolling circle amplification, strand displacement amplification, helicase-dependent amplification), nucleic acid hybridization and/or nucleotide sequencing as are well known in the art. 25 Suitable primers and probes may readily be designed according to the nucleotide sequence set forth in SEQ ID NO:2.

It will also be appreciated that detection may be performed using a composition comprising the isolated protein or fragment thereof, nucleic acids (*e.g.* primers and/or probes), antibodies or antibody fragments that bind the isolated protein in combination 30 with one or more molecules or reagents that facilitate detection. These one or more molecules or reagents may include enzymes, enzyme substrates, DNA polymerases (*e.g.* for PCR detection), secondary antibodies (which may be labeled such as hereinbefore described), buffers and blocking agents, although without limitation thereto. In some embodiments, the invention may provide a detection kit comprising, the isolated

protein, nucleic acids (*e.g.* primers and/or probes), antibodies, antibody fragments, secondary antibodies., the one or more molecules or reagents that facilitate detection and/or compositions comprising same, although without limitation thereto.

As generally used herein, animals may include any species of the animal kingdom susceptible to *E. coli* infections inclusive of humans and non-human mammals and avians. Non-human mammals may include domestic pets (*e.g.* cats, dogs), performance animals (*e.g.* racehorses, camels) and livestock (*e.g.* cattle, pigs, sheep), although without limitation thereto. Avians may include commercially important poultry such as chickens, turkeys and ducks.

So that the invention may be readily understood and put into practical effect, reference is made to the following non-limiting examples.

EXAMPLE 1

MATERIALS AND METHODS

15

The *E. coli* dataset and MLST analysis. The *E. coli* database was represented by 62 complete and 1638 draft genomes available on the NCBI public database as of the 1st January, 2014 (Table 3 provides a listing of 62 complete genome strains).

20 **Bioinformatic analysis.** Sequence comparisons were performed using FASTA36 package (38). Subcellular localization was predicted by PSORTb 3.0 (39) and signal sequence was predicted by LipoP 1.0 (40). Structural analysis was performed by PHYRE2 (41). Prevalence of genes was determined by amino acid sequence identity, using a cut-off of >75% over a 75% alignment. Phylogenetic trees were drawn using
25 MEGA6 (42). MLST analysis was conducted following the scheme previously described (8). Circular representations were drawn using Circos (43).

Bacteria and growth conditions. *E. coli* strains were routinely grown at 37°C on solid or in liquid Luria-Bertani (LB) medium supplemented with appropriate antibiotics;
30 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 minutes at 4°C to generate the whole cell lysate sample. Supernatant fractions

were generated as previously described (44). To obtain EDTA-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=0.5. Cells were centrifuged at 10,000 g for 10 minutes 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 FIG. 5. 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 (45). DNA was transformed into chemically competent *E. coli* TOP10 (Invitrogen) for plasmid propagation. Primers used for cloning are described in FIG. 5. Isolation of plasmid DNA was carried out using the QIAprep Spin Miniprep Kit (Qiagen) and sequenced using Big Dye 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 (Figure 5) as previously described (46). The formation of OMVs was induced by incubation in 0.01M EDTA at 56°C and samples were analyzed by HPLC-MS/MS as previously described (37, 44). 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 (47) 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 (48).

Expression and purification of vaccine targets. Polyhistidine-tagged recombinant proteins were obtained by auto-induction (49) and purified by affinity and gel filtration chromatography. Briefly, BL21 (DE3) pLysS harboring the LIC vectors were grown in ZYP-5052 (1 L) overnight at 28°C, 250 rpm. Cells were harvested by centrifugation at 18,600 x g, resuspended in 25 mM Tris-HCl pH 7, 150 mM NaCl, 0.5% Triton X-100, Protease Inhibitor Cocktail (Sigma) and DNaseI (Roche) and lysed by sonication (Misonix XL-2000, QSonica). Cell debris was removed by centrifugation and histidine-tagged proteins were purified using Talon Metal Affinity resin (Clontech) pre-

equilibrated 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 (ÄKTA, GE Healthcare) using a HiLoad 16/60 Superdex 75 prep grade column (GE Healthcare) pre-equilibrated 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 µg/ml) were coated onto Nunc MaxiSorp flat-bottom 96 well plates (Thermo Scientific) in carbonate coating buffer (18 mM Na₂CO₃, 450 mM NaHCO₃, pH 9.3) at 4°C overnight. Plates were then washed twice with 0.05% Tween20 PBS (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 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 secondary antibody and incubated for 90 min at 37°C. Plates were washed four times with PBST before 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 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® 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 (26), was used to assess the protective efficacy of YncE as a vaccine

immunogen against systemic infection. Groups of twelve C57BL/6 mice (8-12 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 (50). 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 post-challenge by quantitating the bacterial loads in liver, blood, kidney, and spleen. The experiment included mock-immunized controls (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 g for 10 min. Plates were coated, blocked and washed as detailed above. Mice serum samples were applied at a 1:2 serial dilution, starting from a 1:10 dilution. Subsequent wash, secondary antibody and developing steps were performed as previously described (47), with the exception that peroxidase-conjugated anti-mouse antibodies were applied as the secondary antibody. YncE IgG titres were defined as the logarithmic dilution that produced a significant absorbance (450 nm) in comparison to a blank sample.

RESULTS

EcoDS: a comprehensive dataset of *E. coli* genome sequenced strains. An *E. coli* dataset (EcoDS) was generated from 1700 genome sequences available on the NCBI public database (Strains listed in Table 2). EcoDS contained 62 complete (Table 3) and 1638 draft genome sequences, and represents the most diverse composition of *E. coli* strains examined to-date (Tables 1, 2 and Fig. 3). Information regarding the source, disease, origin or year of isolation was available for approximately 69% (n=1174) 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. In order to determine the relationship of the *E. coli* strains in EcoDS, a phylogenetic tree was constructed based on 435 unique multi-locus sequence types (MLST) (8) identified in EcoDS. To determine the phylogenetic group of the *E. coli* strains in EcoDS, we used an *in silico* triplex analysis of the *chuA*, *yjaA* and TSPE4.C2 loci, which allowed the classification of strains into the major phylogroups A, B1, B2 and D (11, 12). This strategy revealed a strong correlation between *E. coli* ST's and the established multi-locus enzyme electrophoresis (MLEE)-based *E. coli* phylogeny, and described a comprehensive distribution of ST's within the four major phylogroups in EcoDS.

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 (13).

EcoDS contained a highly diverse combination of ST's and pathotypes. 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, 14-16) and strains from progenitor EHEC and EAEC ST's 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 EHEC strains involved in UTI cases (17). Moreover, this group also contains non-pathogenic 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) (18, 19). 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 (20, 21). Phylogroup E is the least diverse 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 CE10 and the UPEC strain IAI39. The cryptic lineages C-I to C-V include EHEC, ETEC, commensal and the environmental isolates previously described (20, 21). Moreover, phylogroup D is also represented by strains from ST69 (n=80), which includes the UPEC reference strain UMN026, and the EAEC strain 042. 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 (22, 23) to define a set of 362 *E. coli* essential genes (). We then screened for the prevalence of these essential genes in the 62 complete genomes (Table 3), 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 1700 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, Table 4), leaving a total of 1556 strains in EcoDS (*i.e.* Table 2 strains not in Table 4). The least prevalent essential gene

in EcoDS was *b3119 (tdcR)*, which was present in 99% (n=1541) of strains ().

We used the 1556 genome sequences in EcoDS to define the conserved set of *E. coli* genes. The genome of the well-characterised K-12 strain MG1655 was used as a reference (24) and, based on the essential gene data, a cut-off value for gene prevalence was set at 99%. Pairwise comparison of the 4319 ORFs defined in MG1655 with the genome sequence of the 1556 strains in EcoDS led to the identification of 3042 genes present in more than 99% of strains (protein identity >75% over a 75% sequence overlap), of which 1037 genes were present in 100% of the strains. These 3042 genes define a conserved subset of *E. coli* genes in the 1556 strains that make up EcoDS .

In order to define the *E. coli* accessory genome, 298,563 annotated ORFs (which comprised 290,776 chromosomal ORFs and 7787 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, 2209 plasmid), and 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.

20 Identification of novel vaccine targets. The definition of a conserved set of 3042 *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 included in a broadly protective vaccine against *E. coli*, we examined the surface-associated outer membrane proteome of the MDR ST131 strain EC958. Proteomic analysis of outer membrane vesicles (OMVs) induced from EC958 led to the identification of 115 proteins. 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 N-terminal 6xHis-tag sequence. Expression studies using these constructs revealed only the BamC and YncE proteins were produced as soluble recombinant proteins (Fig. 4A). 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 SslE (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 (25). Both SslE ($P<0.001$) and EsiB ($P=0.022$) showed higher reactivity to the plasma from urosepsis patients compared to healthy individuals (Fig. 1A). Among the targets 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 ($P<0.001$), 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 (26). YncE was highly immunogenic inducing a strong IgG response (Fig. 4B) and mice immunized with YncE were significantly protected against infection as evidenced by lower blood and liver *E. coli* loads following systemic challenge (Fig. 1B). 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 ($n=1550/1556$, 99.6%). Strains lacking *yncE* gene were not related by ST or phylogroup: DEC13A, DEC13B, DEC13D, IAI39, N1, UMEA 3687-1. 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. 2A). It was observed that YncE is broadly expressed by *E. coli* strains from different

pathotypes. Moreover, western blot analysis of strains representing the *E. coli* phylogroups determined in this study revealed the presence of YncE in the supernatant after overnight growth (Fig. 2A). YncE was also overexpressed in an *E. coli fur* mutant during *in vitro* growth, demonstrating a role for iron in its regulatory control (Fig. 2B).

5

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 (27), components and effectors of the type III secretion system from EPEC (28), Shiga toxin from EHEC (29) and fimbrial adhesins and siderophore receptors from UPEC (30). In this study we used an essential gene strategy to generate a curated dataset of *E. coli* genomes (EcoDS), and used this comprehensive and diverse collection (representing 435 STs) to define core and accessory elements of the pan *E. coli* genome. 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 housekeeping MLST genes was consistent with a modern framework composed of phylogroups A, AxB1, B2, and D. 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 (31). In this study, we observed that the *E. coli* accessory genome comprises a broad set of 12,722 ORFs (prevalence <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 SslE, for example, were present in approximately 89% and 70% of the strains in EcoDS,

respectively.

A smaller subset of 3042 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. 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 SslE, a type II secreted-mucinase originally identified from ExPEC (25) that provides broad protection against *E. coli* infection in different animal models (32). BamC, a surface-associated lipoprotein (33), 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 (34) transported by the Sec machinery (35) and associated with binding to single stranded DNA (36). We have previously shown that YncE is present in the OMV proteome of a large collection of urosepsis strains (37) 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, 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 non-pathogenic *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.

Throughout the specification the aim has been to describe the preferred embodiments of the invention without limiting the invention to any one embodiment or specific collection of features. It will therefore be appreciated by those of skill in the art that, in light of the instant disclosure, various modifications and changes can be made in
5 the particular embodiments exemplified without departing from the scope of the present invention.

All computer programs, algorithms, patent and scientific literature referred to herein is incorporated herein by reference.

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

Source	%	Disease	%	Year	%	Place	
Bat	0.1	Asymptomatic	0.8	<1940s	0.1	Africa	0
Buffalo	0.1	Bacteremia	13.1	1950s	0.5	Asia	1
Cow	4.5	Bacteriuria	0.1	1960s	0.6	Europe	0
Dog	0.2	Crohn's disease	0.2	1970s	0.7	North America	1
Environment	0.9	Diarrhea	6.7	1980s	2.2	South America	0
Fish	0.1	HUS	1.1	1990s	5.5	Oceania	0
Food	1.0	Mastitis	0.4	2000s	12.2	Unknown	5
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						

RTI, respiratory tract infections; UTI, urinary tract infections; HUS, hemolytic-uremic syndrome

Table 2. 1700 *E. coli* strains used in this study

BL21-Gold; 042; 536; 55989; ABU 83972; APEC O1; APEC O78; ATCC 8739; REL606; BL21; BW2952; c321.deltaA; CFT073; DH1; DH1 (ME8569); E24377A; ED1a; H10407; HS; IAI1; IAI39; IHE3034; JJ1886; KO11FL; LF82; LY180; NA114; 12009; 2009EL-2050; 2009EL-2071; 2011C-3493; 11128; E2348/69; EC4115; EDL933; Sakai; TW14359; EC958; 11368; CB9615; RM12579; CE10; NRG857c; P12b; PMV-1; S88; SE11; SE15; SMS-3-5; clone D i14; clone D i2; DH10B; MC4100; MDS42; MG1655; W3110; UM146; UMN026; UMN18; UMNK88; UTI89; Xuzhou21; 0.1288; 0.1304; 3.4870; 3.4880; 5.2239; 6.0172; 7.1982; 8.0416; 8.0566; 8.0569; 8.0586; 8.2524; 10.0821; 10.0833; 10.0869; 38.16; 38.27; 38.34; 38.52; 75; 88.0221; 88.1042; 88.1467; 89.0511; 90.0039; 90.0091; 90.2281; 93.0055; 93.0056; 94.0618; 95.0083; 95.0183; 95.0943; 95.1288; 96.0107; 96.0109; 96.0427; 96.0428; 96.0932; 96.0939; 97.0003; 97.0007; 97.0010; 97.1742; 99.067; 99.0672; 99.0678; 99.0713; 99.0814; 99.0815; 99.0816; 99.0839; 99.0848; 99.1753; 99.1762; 99.1775; 99.1781; 99.1793; 99.1805; 148; 320; 328; 435; 597; 606; 668; 681; 897; 1044; 1047; 1125; 1180; 1240; 1350; 1357; 1365; 3003; 3006; 3431; 5412; 5905; 07798; 8624; 1.2264; 1.2741; 2.3916; 2.4168; 3.2303; 3.2608; 3.3884; 4.0522; 4.0967; 5.0588; 5.0959; 53638; 83972; 9.0111; 9.1649; 96.154; 110957; 113290; 113302; 113303; 174750; 174900; 178200; 178850; 178900; 179100; 179550; 180050; 180200; 180600; 199900.1; 201600.1; 11-3677; 900105; 907357; 907391; 907446; 907672; 907700; 907701; 907710; 907713; 907715; 907779; 907889; 907892; 908519; 908521; 908522; 908524; 908525; 908541; 908555; 908573; 908585; 908616; 908624; 908632; 908658; 908675; 908691; 909957; 11-4404; 93.0624; 95.0941; 11-4522; 96.0497; 97.0246; 97.0259; 97.0264; 99.0741; 11-4632; 09-7901; 04-8351; 2719100; 2720900; 2722950; 2726800; 2726950; 2729250; 2730350; 2730450; 2731150; 2733950; 2735000; 2741950; 2747800; 2749250; 2756500; 2762100; 2770900; 2780750; 2785200; 2788150; 2845350; 2845650; 2846750; 2848050; 2850400; 2850750; 2851500; 2853500; 2854350; 2860050; 2860650; 2861200; 2862600; 2864350; 2865200; 2866350; 2866450; 2866550; 2866750; 2867750; 2871950; 2872000; 2872800; 2875000; 2875150; 112469215; 112469218; 01-09591; 03-EN-705; 08BKT055439; 08BKT77219; 09BKT024447; 09BKT076207; 09BKT078844; 101-1; 11-02030; 11-02033-1; 11-02092; 11-02093; 11-02281; 11-02318; 11-02913; 11-03439; 11-03943; 11-04080; 11-4632 C1; 11-4632 C2; 11-4632 C3; 11-4632 C4; 11-4632 C5; 14A; 1827-70; 1A; 2362-75; 2534-86; 27A; 2886-75; 3030-1; 3256-97; 4_1_47FAA; 4865/96; 493/89; 53A; 53C; 541-1; 541-15; 576-1; 85B; 87A; 909945-2; 910096-2; 93-001; 95JB1; 95NR1; 9B; A25922R; A35218R; AA86; AB42410445; AB42554418; AB42602061; AB43739056; AD30; AI27; ARS4.2123; ATCC 23502; ATCC 23506; ATCC 25922; ATCC 35150; ATCC 700728; B088; B093; B102; B103; B104; B105; B106; B107; B108; B109; B112; B113; B114; B15; B17; B171; B175; B185; B26-1; B26-2; B28-1; B28-2; B29-1; B29-2; B2F1; B354; B36-1; B36-2; B367; B40; B40-1; B40-2; B41; B49-2; B5-2; B574; B671; B7-1; B7-2; B706; B799; B7A; B83; B84; B85; B86; B89; B90; B921; B93; B94; B95; BAA-2192; BAA-2193; BAA-2196; BAA-2209; BAA-2215; BAA-2219; BCE001_MS16; BCE002_MS12; BCE006_MS-23; BCE007_MS-11; BCE008_MS-01; BCE008_MS-13; BCE011_MS-01; BCE019_MS-13; BCE030_MS-09; BCE032_MS-12; BCE034_MS-14; Bd5610_99; BIDMC 19C; BIDMC 37; BIDMC 38; BIDMC 39; BWH 24; BWH 32; C-34666; C12_92; C1214_90; C1244_91; C154_11; C155_11; C157_11; C161_11; C166_11; C170_11; C213_10; C2139_99; C227-11; C236-11; C238_91; C260_92; C262_10; C283_09; C295_10; C341_10; C343_08; C347_93; C353_09; C354_03B; C40_11; C418_89; C43/90; C458_10; C48/93; C488_07; C496_10; C497_10; C527_94; C58_11; C581_05; C586_05; C639_08; C652_10; C654_09; C717_10; C725_88; C732_98; C743_03; C751_03; C78_09C; C79_08; C792_92; C796_10; C799_92; C80_08; C807_09; C82_11; C824_10; C842_97; C844-97; C87_11; C887_10; C9_92; C900_01; C93_11; CB7326; CFSAN001629; CFSAN001630; CFSAN001632; CFSAN002236; CFSAN002237; chi7122; CL-3; clone A i1; CUMT8; CVDNalr; CVM10021; CVM10026; CVM10030; CVM10224; CVM9340; CVM9450; CVM9455; CVM9534; CVM9545; CVM9553; CVM9570; CVM9574; CVM9602; CVM9634; CVM9942; CVM9952; DEC10A; DEC10B; DEC10C; DEC10D; DEC10E; DEC10F; DEC11A; DEC11B; DEC11C; DEC11D; DEC11E; DEC12A; DEC12B; DEC12C; DEC12D; DEC12E; DEC13A; DEC13B; DEC13C; DEC13D; DEC13E; DEC14A; DEC14B; DEC14C; DEC14D; DEC15A; DEC15B; DEC15C; DEC15D; DEC15E; DEC1A; DEC1B; DEC1C; DEC1D; DEC1E; DEC2A; DEC2B; DEC2C; DEC2D; DEC2E; DEC3A; DEC3B; DEC3C; DEC3D; DEC3E; DEC3F; DEC4A; DEC4B; DEC4C; DEC4D; DEC4E; DEC4F; DEC5A; DEC5B; DEC5C; DEC5D; DEC5E; DEC6A; DEC6B; DEC6C; DEC6D; DEC6E; DEC7A; DEC7B; DEC7C; DEC7D; DEC7E; DEC8A; DEC8B; DEC8C; DEC8D; DEC8E; DEC9A; DEC9B; DEC9C; DEC9D; DEC9E; DORA_A_5_14_21; DORA_B_14; E1002; E101; E110019; E1114; E1118; E112/10; E1167; E128010; E1492; E1520; E1777; E22; E2265; E267; E482; E560; E704; E92/11; Ec11-4984; Ec11-4986; Ec11-4987; Ec11-4988; Ec11-5536; Ec11-5537; Ec11-5538; Ec11-5603; Ec11-5604; Ec11-6006; Ec11-9450; Ec11-9941; Ec11-9990; Ec12-0465;

Ec12-0466; EC1212; EC1734; EC1735; EC1736; EC1737; EC1738; EC1845; EC1846; EC1847; EC1848; EC1849; EC1850; EC1856; EC1862; EC1863; EC1864; EC1865; EC1866; EC1868; EC1869; EC1870; EC302/04; EC4009; EC4013; EC4024; EC4042; EC4045; EC4076; EC4084; EC4100B; EC4113; EC4127; EC4191; EC4192; EC4196; EC4203; EC4205; EC4206; EC4401; EC4402; EC4421; EC4422; EC4436; EC4437; EC4439; EC4448; EC4486; EC4501; EC508; EC536; EC869; EC96038; ECA-0157; ECA-727; ECC-1470; ECC-Z; Envira 10/1; Envira 8/11; EPEC C342-62; EPECa12; EPECa14; F11; FAH1; FAH2; FAP1; FAP2; FBH1; FBP1; FCH1; FCP1; FDA504; FDA505; FDA506; FDA507; FDA517; FRIK1985; FRIK1990; FRIK1996; FRIK1997; FRIK1999; FRIK2000; FRIK2001; FRIK523; FRIK920; FRIK966; FVEC 1465; FVEC1302; FVEC1412; G5101; G58-1; GOS1; GOS2; GSK2022; GSK2023; GSK2024; GSK202B; GSK202U; H001; H093800014; H112180280; H112180282; H112180283; H112180540; H112180541; H120; H185; H218; H220; H223; H252; H263; H2687; H288; H296; H299; H30; H305; H378; H383; H386; H397; H413; H420; H442; H454; H461; H489; H494; H504; H588; H591; H593; H605; H617; H660; H730; H736; HM26; HM27; HM46; HM605; HM65; HM69; HVH 1; HVH 10; HVH 100; HVH 102; HVH 103; HVH 104; HVH 106; HVH 107; HVH 108; HVH 109; HVH 110; HVH 111; HVH 112; HVH 113; HVH 114; HVH 115 (4-4465989); HVH 115 (4-4465997); HVH 116; HVH 117; HVH 118; HVH 119; HVH 12; HVH 120; HVH 121; HVH 122; HVH 125; HVH 126; HVH 127; HVH 128; HVH 13; HVH 130; HVH 132; HVH 133; HVH 134; HVH 135; HVH 136; HVH 137; HVH 138; HVH 139; HVH 140; HVH 141; HVH 142; HVH 143; HVH 144; HVH 145; HVH 146; HVH 147; HVH 148; HVH 149; HVH 150; HVH 151; HVH 152; HVH 153; HVH 154; HVH 155; HVH 156; HVH 157; HVH 158; HVH 159; HVH 16; HVH 160; HVH 161; HVH 162; HVH 163; HVH 164; HVH 167; HVH 169; HVH 17; HVH 170; HVH 171; HVH 172; HVH 173; HVH 175; HVH 176; HVH 177; HVH 178; HVH 18; HVH 180; HVH 182; HVH 183; HVH 184; HVH 185; HVH 186; HVH 187; HVH 188; HVH 189; HVH 19; HVH 190; HVH 191; HVH 192; HVH 193; HVH 194; HVH 195; HVH 196; HVH 197; HVH 198; HVH 199; HVH 2; HVH 20; HVH 200; HVH 201; HVH 202; HVH 203; HVH 204; HVH 205; HVH 206; HVH 207; HVH 208; HVH 209; HVH 21; HVH 210; HVH 211; HVH 212; HVH 213; HVH 214; HVH 215; HVH 216; HVH 217; HVH 218; HVH 22; HVH 220; HVH 221; HVH 222; HVH 223; HVH 225; HVH 227; HVH 228; HVH 23; HVH 24; HVH 25; HVH 26; HVH 27; HVH 28; HVH 29; HVH 3; HVH 30; HVH 31; HVH 32; HVH 33; HVH 35; HVH 36; HVH 37; HVH 38; HVH 39; HVH 4; HVH 40; HVH 41; HVH 42; HVH 43; HVH 44; HVH 45; HVH 46; HVH 48; HVH 5; HVH 50; HVH 51; HVH 53; HVH 55; HVH 56; HVH 58; HVH 59; HVH 6; HVH 61; HVH 63; HVH 65; HVH 68; HVH 69; HVH 7; HVH 70; HVH 73; HVH 74; HVH 76; HVH 77; HVH 78; HVH 79; HVH 80; HVH 82; HVH 83; HVH 84; HVH 85; HVH 86; HVH 87; HVH 88; HVH 89; HVH 9; HVH 90; HVH 91; HVH 92; HVH 95; HVH 96; HVH 98; IMT8073; J53; J96; JB1-95; Jurua 18/11; Jurua 20/10; K1; K2; KD1; KD2; KOEGE 10; KOEGE 118; KOEGE 131; KOEGE 3; KOEGE 30; KOEGE 32; KOEGE 33; KOEGE 40; KOEGE 43; KOEGE 44; KOEGE 56; KOEGE 58; KOEGE 61; KOEGE 62; KOEGE 68; KOEGE 7; KOEGE 70; KOEGE 71; KOEGE 73; KOEGE 77; KTE1; KTE10; KTE100; KTE101; KTE102; KTE103; KTE104; KTE105; KTE106; KTE107; KTE108; KTE109; KTE11; KTE111; KTE112; KTE113; KTE114; KTE115; KTE116; KTE117; KTE118; KTE119; KTE12; KTE120; KTE121; KTE122; KTE123; KTE124; KTE125; KTE126; KTE127; KTE128; KTE129; KTE13; KTE130; KTE131; KTE132; KTE133; KTE134; KTE135; KTE136; KTE137; KTE138; KTE139; KTE14; KTE140; KTE141; KTE142; KTE143; KTE144; KTE145; KTE146; KTE147; KTE148; KTE15; KTE150; KTE153; KTE154; KTE155; KTE156; KTE157; KTE158; KTE159; KTE16; KTE160; KTE161; KTE162; KTE163; KTE165; KTE166; KTE167; KTE168; KTE169; KTE17; KTE170; KTE171; KTE172; KTE173; KTE174; KTE175; KTE176; KTE177; KTE178; KTE179; KTE18; KTE180; KTE181; KTE182; KTE183; KTE184; KTE185; KTE186; KTE187; KTE188; KTE189; KTE19; KTE190; KTE191; KTE192; KTE193; KTE194; KTE195; KTE196; KTE197; KTE198; KTE199; KTE2; KTE20; KTE200; KTE201; KTE202; KTE203; KTE204; KTE205; KTE206; KTE207; KTE208; KTE209; KTE21; KTE210; KTE211; KTE212; KTE213; KTE214; KTE215; KTE216; KTE217; KTE218; KTE219; KTE22; KTE220; KTE221; KTE222; KTE223; KTE224; KTE225; KTE226; KTE227; KTE228; KTE229; KTE23; KTE230; KTE231; KTE232; KTE233; KTE234; KTE235; KTE236; KTE237; KTE24; KTE240; KTE25; KTE26; KTE27; KTE28; KTE29; KTE3; KTE31; KTE33; KTE34; KTE35; KTE36; KTE37; KTE38; KTE39; KTE4; KTE40; KTE41; KTE42; KTE43; KTE44; KTE45; KTE46; KTE47; KTE48; KTE49; KTE5; KTE50; KTE51; KTE52; KTE53; KTE54; KTE55; KTE56; KTE57; KTE58; KTE59; KTE6; KTE60; KTE61; KTE62; KTE63; KTE64; KTE65; KTE66; KTE67; KTE68; KTE69; KTE7; KTE70; KTE71; KTE72; KTE73; KTE74; KTE75; KTE76; KTE77; KTE78; KTE79; KTE8; KTE80; KTE81; KTE82; KTE83; KTE84; KTE85; KTE86; KTE87; KTE88; KTE89; KTE9; KTE90; KTE91; KTE93; KTE94; KTE95; KTE96; KTE97; KTE98; KTE99; LAU-EC1; LAU-EC10; LAU-EC2; LAU-EC3; LAU-EC4; LAU-EC5; LAU-EC6; LAU-EC7; LAU-EC8; LAU-EC9; LB226692; LCT-EC106; LCT-EC52; LCT-EC59; LSU-61; LT-68; M056; M1; M10; M11; M114; M12; M13; M14; M15;

M16; M17; M18; M19; M2; M20; M21; M22; M23; M3; M4; M5; M6; M605; M646; M7; M718; M8; M863; M9; M919; MA6; MC19; MC21; MC23; MC6002; MC6003; MG1655star; MP020940.1; MP020980.1; MP020980.2; MP021017.1; MP021017.10; MP021017.11; MP021017.12; MP021017.2; MP021017.3; MP021017.4; MP021017.5; MP021017.6; MP021017.9; MP021552.11; MP021552.12; MP021552.7; MP021552.8; MP021561.2; MP021561.3; MP021566.1; MS 107-1; MS 110-3; MS 115-1; MS 116-1; MS 117-3; MS 119-7; MS 124-1; MS 145-7; MS 146-1; MS 153-1; MS 16-3; MS 175-1; MS 182-1; MS 185-1; MS 187-1; MS 196-1; MS 198-1; MS 200-1; MS 21-1; MS 45-1; MS 57-2; MS 60-1; MS 69-1; MS 78-1; MS 79-10; MS 84-1; MS 85-1; MT#2; N1; NC101; NCCP15647; NCCP15657; NCCP15658; NCCP15738; NCCP15739; NE037; NE098; NE1487; NIPH-11060424; Nissle 1917; O08; O91; OK1114; OK1180; OK1357; ON2010; ON2011; OP50; P0298942.1; P0298942.10; P0298942.11; P0298942.12; P0298942.14; P0298942.15; P0298942.2; P0298942.3; P0298942.4; P0298942.6; P0298942.7; P0298942.8; P0298942.9; P0299438.10; P0299438.11; P0299438.2; P0299438.3; P0299438.4; P0299438.5; P0299438.6; P0299438.7; P0299438.8; P0299438.9; P0299483.1; P0299483.2; P0299483.3; P02997067.6; P0299917.1; P0299917.10; P0299917.2; P0299917.3; P0299917.4; P0299917.5; P0299917.6; P0299917.7; P0299917.8; P0299917.9; P0301867.1; P0301867.11; P0301867.13; P0301867.2; P0301867.3; P0301867.4; P0301867.5; P0301867.7; P0301867.8; P0301904.3; P0302293.10; P0302293.2; P0302293.3; P0302293.4; P0302293.6; P0302293.7; P0302293.8; P0302293.9; P0302308.1; P0302308.10; P0302308.11; P0302308.12; P0302308.13; P0302308.14; P0302308.2; P0302308.3; P0302308.4; P0302308.5; P0304777.1; P0304777.10; P0304777.11; P0304777.12; P0304777.13; P0304777.14; P0304777.15; P0304777.2; P0304777.3; P0304777.4; P0304777.5; P0304777.7; P0304777.8; P0304777.9; P0304799.3; P0304816.1; P0304816.10; P0304816.11; P0304816.12; P0304816.13; P0304816.14; P0304816.15; P0304816.2; P0304816.3; P0304816.4; P0304816.5; P0304816.6; P0304816.7; P0304816.8; P0304816.9; P0305260.1; P0305260.10; P0305260.11; P0305260.12; P0305260.13; P0305260.15; P0305260.2; P0305260.3; P0305260.4; P0305260.5; P0305260.6; P0305260.7; P0305260.8; P0305260.9; p0305293.1; p0305293.10; p0305293.11; p0305293.12; p0305293.13; p0305293.14; p0305293.15; p0305293.2; p0305293.3; p0305293.4; p0305293.5; p0305293.6; p0305293.7; p0305293.8; p0305293.9; P4; P4-96; P4-NR; PA10; PA11; PA13; PA14; PA15; PA19; PA2; PA22; PA23; PA24; PA25; PA28; PA3; PA31; PA32; PA33; PA34; PA35; PA38; PA39; PA4; PA40; PA41; PA42; PA45; PA47; PA48; PA49; PA5; PA7; PA8; PA9; PCN033; PUTI459; R424; R527; R529; RN587/1; S17; SCD1; SCD2; SCI-07; SEPT362; STEC_7v; STEC_94C; STEC_B2F1; STEC_C165-02; STEC_DG131-3; STEC_EH250; STEC_H.1.8; STEC_MHI813; STEC_O31; STEC_S1191; SWW33; T1282_01; T1840_97; T22; T234_00; T408; T426; T924_01; TA004; TA007; TA008; TA014; TA024; TA054; TA103; TA124; TA141; TA143; TA144; TA206; TA249; TA255; TA271; TA280; TA435; TA447; TA464; ThroopD; TOP2386; TOP2396-1; TOP2396-2; TOP2396-3; TOP2515; TOP2522-1; TOP2652; TOP2662-1; TOP2662-2; TOP2662-3; TOP2662-4; TOP291; TOP293-1; TOP293-2; TOP293-3; TOP293-4; TOP379; TOP382-1; TOP382-2; TOP382-3; TOP498; TOP550-1; TOP550-2; TOP550-3; TOP550-4; TT12B; TTSH84137; TW00353; TW06591; TW07509; TW07793; TW07945; TW09098; TW09109; TW09195; TW09231; TW09276; TW10119; TW10246; TW10509; TW10598; TW10722; TW10828; TW11039; TW11681; TW14182; TW14301; TW14313; TW14425; TW14588; TW15838; TW15901; Tx1686; TX1999; Tx3800; TY-2482; UMD753; UMEA 3014-1; UMEA 3022-1; UMEA 3033-1; UMEA 3041-1; UMEA 3052-1; UMEA 3053-1; UMEA 3065-1; UMEA 3087-1; UMEA 3088-1; UMEA 3097-1; UMEA 3108-1; UMEA 3113-1; UMEA 3117-1; UMEA 3121-1; UMEA 3122-1; UMEA 3124-1; UMEA 3139-1; UMEA 3140-1; UMEA 3144-1; UMEA 3148-1; UMEA 3150-1; UMEA 3151-1; UMEA 3152-1; UMEA 3155-1; UMEA 3159-1; UMEA 3160-1; UMEA 3161-1; UMEA 3162-1; UMEA 3163-1; UMEA 3172-1; UMEA 3173-1; UMEA 3174-1; UMEA 3175-1; UMEA 3176-1; UMEA 3178-1; UMEA 3180-1; UMEA 3185-1; UMEA 3190-1; UMEA 3193-1; UMEA 3199-1; UMEA 3200-1; UMEA 3201-1; UMEA 3203-1; UMEA 3206-1; UMEA 3208-1; UMEA 3212-1; UMEA 3215-1; UMEA 3216-1; UMEA 3217-1; UMEA 3220-1; UMEA 3221-1; UMEA 3222-1; UMEA 3230-1; UMEA 3233-1; UMEA 3240-1; UMEA 3244-1; UMEA 3257-1; UMEA 3264-1; UMEA 3268-1; UMEA 3271-1; UMEA 3290-1; UMEA 3292-1; UMEA 3298-1; UMEA 3304-1; UMEA 3314-1; UMEA 3317-1; UMEA 3318-1; UMEA 3323-1; UMEA 3329-1; UMEA 3336-1; UMEA 3337-1; UMEA 3341-1; UMEA 3342-1; UMEA 3355-1; UMEA 3391-1; UMEA 3426-1; UMEA 3489-1; UMEA 3490-1; UMEA 3585-1; UMEA 3592-1; UMEA 3609-1; UMEA 3617-1; UMEA 3632-1; UMEA 3652-1; UMEA 3656-1; UMEA 3662-1; UMEA 3671-1; UMEA 3682-1; UMEA 3687-1; UMEA 3693-1; UMEA 3694-1; UMEA 3702-1; UMEA 3703-1; UMEA 3705-1; UMEA 3707-1; UMEA 3718-1; UMEA 3805-1; UMEA 3821-1; UMEA 3834-1; UMEA 3889-1; UMEA 3893-1; UMEA 3899-1; UMEA 3955-1; UMEA 4075-1; UMEA 4076-1; UMEA 4207-1; USDA 5905; W; W26; Wa1; Wa2; WC1; WC2; WV_060327; XH001; XH140A;

Table 3. 62 completely sequenced *E. coli* strains used in this study

BL21-Gold; 042; 536; 55989; ABU 83972; APEC O1; APEC O78; ATCC 8739; REL606; BL21; BW2952; c321.deltaA; CFT073; DH1; DH1 (ME8569); E24377A; ED1a; H10407; HS; IAI1; IAI39; IHE3034; JJ1886; KO11FL; LF82; LY180; NA114; 12009; 2009EL-2050; 2009EL-2071; 2011C-3493; 11128; E2348/69; EC4115; EDL933; Sakai; TW14359; EC958; 11368; CB9615; RM12579; CE10; NRG857c; P12b; PMV-1; S88; SE11; SE15; SMS-3-5; clone D i14; clone D i2; DH10B; MC4100; MDS42; MG1655; W3110; UM146; UMN026; UMN18; UMNK88; UTI89; Xuzhou21.

Table 4. 144 *E. coli* strains missing more than one essential gene

08bkt77219; 09bkt024447; 101-1; 1044; 1125; 174900; 178850; 180050; 1a; 2722950; 27a; 2862600; 2872800; 3.4880; 541-1; 87a; 908519; 909957; 910096-2; 95.0083; 95jb1; 97.1742; 99.1775; ai27; b114; b171; b40; b7a; b94; bce008_ms-01; c1244_91; c155_11; c58_11; clonea_i1; cvm10021; cvm9455; cvm9534; cvm9545; cvm9553; cvm9570; cvm9574; cvm9952; dec10f; dec11e; dec12a; dec1b; dec2e; dec3c; dec4a; dec7b; dec9c; dora_a_5_14_21; dora_b_14; e110019; e22; e482; ec4024; ec4076; ec4113; ec4192; ec4205; ec4401; ec4421; ec4448; ec4486; ec4501; ec508; ec869; eca-0157; eca-727; ecc-1470; f11; gsk202b; h001; h093800014; h112180283; h112180540; h112180541; h494; h591; hm27; hm69; hvh_221; hvh_222; hvh_225; hvh_29; k1; koege_61; koege_68; koege_71; kte103; kte204; kte234; kte39; kte70; kte90; lau-ec1; m13; m1; m21; m22; m23; m5; m8; m919; mp021017.1; ms_124-1; op50; p0298942.7; p0301867.8; p0304816.14; p0304816.1; p0305293.13; pa19; scd1; scd2; sci-07; ta004; ta014; ta141; ta206; top2386; top2396-3; top2515; top2522-1; top291; top293-1; top293-3; top293-4; top382-2; top498; top550-1; ttsh84137; tw09231; tw09276; tw10722; tw11039; tw14182; tw15838; umea_3257-1; umea_3298-1; umea_3899-1; wa2; wv_060327.

Table 5. *E. coli* strains missing YncE gene

DEC13A; DEC13B; DEC13D; IAI39; N1; UMEA 3687-1

REFERENCES

1. **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-Moller 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.
2. **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.
3. **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.
4. **Kaper JB, Nataro JP, Mobley HL.** 2004. Pathogenic *Escherichia coli*. *Nat Rev Microbiol* **2**:123-140.
5. **Didelot X, Meric G, Falush D, Darling AE.** 2012. Impact of homologous and non-homologous recombination in the genomic evolution of *Escherichia coli*. *BMC Genomics* **13**:256.
6. **Jaureguy F, Landreau 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.
7. **von Mentzer A, Connor TR, Wieler LH, Semmler T, Iguchi A, Thomson NR, Rasko DA, Joffre E, Corander J, Pickard D, Wiklund G, Svennerholm AM, Sjoling A, Dougan G.** 2014. Identification of enterotoxigenic *Escherichia coli* (ETEC) clades with long-term global distribution. *Nat Genet* doi:10.1038/ng.3145.

8. **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.
9. **Karch H, Denamur E, Dobrindt U, Finlay BB, Hengge R, Johannes L, Ron EZ, Tonjum 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.
10. **Moriel DG, Rosini R, Seib KL, Serino L, Pizza M, Rappuoli R.** 2012. *Escherichia coli*: Great Diversity around a Common Core. *MBio* **3**.
- 10 11. **Clermont O, Bonacorsi S, Bingen E.** 2000. Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl Environ Microbiol* **66**:4555-4558.
12. **Escobar-Paramo 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.
- 15 13. **Salipante SJ, Roach DJ, Kitzman JO, Snyder MW, Stackhouse B, Butler-Wu SM, Lee C, Cookson BT, Shendure J.** 2014. Large-scale genomic sequencing of extraintestinal pathogenic *Escherichia coli* strains. *Genome Res* doi:10.1101/gr.180190.114.
- 20 14. **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.
- 25 15. **Brzuszkiewicz E, Thurmer 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
- 30

- reveal the emergence of a new pathotype: Entero-Aggregative-Haemorrhagic *Escherichia coli* (EAHEC). *Arch Microbiol* **193**:883-891.
16. **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.
 17. **Toval F, Schiller R, Meisen I, Putze J, Kouzel IU, Zhang W, Karch H, Bielaszewska M, Mormann M, Muthing J, Dobrindt U.** 2014. Characterization of urinary tract infection-associated Shiga toxin-producing *Escherichia coli*. *Infect Immun* **82**:4631-4642.
 18. **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-Bano 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.
 19. **Price LB, Johnson JR, Aziz M, Clabots C, Johnston B, Tchesnokova V, Nordstrom L, Billig M, Chattopadhyay 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**.
 20. **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.
 21. **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.
 22. **Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martinez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M, Latendresse M, Muniz-Rascado L, Ong Q, Paley S, Schroder 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**:D605-612.
23. **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.
24. **Blattner FR, Plunkett G, 3rd, 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**:1453-1474.
25. **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.
26. **Smith SN, Hagan EC, Lane MC, Mobley HL.** 2010. Dissemination and Systemic Colonization of Uropathogenic *Escherichia coli* in a Murine Model of Bacteremia. *MBio* **1**.
27. **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.
28. **Horne C, Vallance BA, Deng W, Finlay BB.** 2002. Current progress in enteropathogenic and enterohemorrhagic *Escherichia coli* vaccines. *Expert Rev Vaccines* **1**:483-493.
29. **Garcia-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.
30. **Moriel DG, Schembri MA.** 2013. Vaccination approaches for the prevention of urinary tract infection. *Curr Pharm Biotechnol* **14**:967-974.
31. **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.

32. Nesta B, Valeri M, Spagnuolo A, Rosini R, Mora M, Donato P, Alteri CJ, Del Vecchio M, Buccato S, Pezzicoli A, Bertoldi I, Buzzigoli L, Tuscano G, Falduto M, Ripa V, Ashhab Y, Bensi G, Fontana MR, Seib KL, Mobley HL, Pizza M, Soriani M, Serino L. 2014. SslE 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.
33. 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.
34. 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.
35. 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.
36. 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.
37. 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* **115C**:93-106.
38. Pearson WR, Wood T, Zhang Z, Miller W. 1997. Comparison of DNA sequences with protein sequences. *Genomics* **46**:24-36.
39. 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.
40. 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.

41. **Kelley LA, Sternberg MJ.** 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**:363-371.
42. **Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S.** 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**:2725-2729.
43. **Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA.** 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* **19**:1639-1645.
44. **Moriel DG, Beatson SA, Wurpel DJ, Lipman J, Nimmo GR, Paterson DL, Schembri MA.** 2013. Identification of Novel Vaccine Candidates against Multidrug-Resistant *Acinetobacter baumannii*. *PLoS One* **8**:e77631.
45. **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.
46. **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.
47. **Nichols KB, Totsika M, Moriel DG, Lo AW, Yang J, Wurpel 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* doi:10.1128/JB.00791-15.
48. **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.
49. **Studier FW.** 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**:207-234.
50. **Brumbaugh AR, Smith SN, Mobley HL.** 2013. Immunization with the Yersiniabactin Receptor, FyuA, Protects Against Pyelonephritis in a Murine Model of Urinary Tract Infection. *Infection and Immunity* **81**:3309-3316.

CLAIMS

1. A method of eliciting an immune response in an animal, said method including the step of administering to the animal: an isolated protein comprising the amino acid sequence of SEQ ID NO:1; one or more immunogenic fragments, variants and/or derivatives thereof; an isolated nucleic acid encoding said isolated protein, immunogenic fragment, variant and/or derivative; and/or an antibody or antibody fragment that binds said isolated protein, immunogenic fragment, variant and/or derivative; to thereby elicit an immune response to *E. coli* in the animal.
2. A method of immunizing an animal including the step of administering to the animal: an isolated protein comprising the amino acid sequence of SEQ ID NO:1; one or more immunogenic fragments, variants and/or derivatives thereof; an isolated nucleic acid encoding said isolated protein, immunogenic fragment, variant and/or derivative; and/or an antibody or antibody fragment that binds said isolated protein, immunogenic fragment, variant and/or derivative; to thereby induce immunity to *E. coli* in the animal.
3. A method of treating an *E. coli*-associated disease, disorder or condition in an animal, said method including the step of administering to the animal: an isolated protein comprising the amino acid sequence of SEQ ID NO:1; one or more immunogenic fragments, variants and/or derivatives thereof; an isolated nucleic acid encoding said isolated protein, immunogenic fragment, variant and/or derivative; and/or an antibody or antibody fragment that binds said isolated protein, immunogenic fragment, variant and/or derivative; to thereby treat the disease, disorder or condition in the animal.
4. The method of Claim 3, wherein the disease, disorder or condition is, or includes, extra-intestinal infections such as urinary tract infection, sepsis and meningitis; and intestinal infections that lead to watery/bloody diarrhea, hemorrhagic colitis and hemolytic uremic syndrome.
5. The method of any preceding claim, wherein the isolated protein, one or more immunogenic fragments, variants and/or derivatives thereof, the isolated nucleic acid and/or the antibody or antibody fragment elicits an immune response to, induces immunity to, or prevents or treats a disease disorder or condition associated with, a plurality of different strains, serotypes or pathotypes of *E.*

coli.

6. A method of detecting an *E. coli* bacterium, or one or more molecular components thereof, in a sample, said method including the step of detecting: an isolated protein comprising the amino acid sequence of SEQ ID NO:1; one or more immunogenic fragments, variants and/or derivatives thereof; an isolated nucleic acid encoding said isolated protein, immunogenic fragment, variant and/or derivative; and/or an antibody or antibody fragment that binds said isolated protein, immunogenic fragment, variant and/or derivative; in said sample, which indicates the presence of *E. coli*, or one or more molecular components thereof, in the sample.
7. The method of any preceding claim wherein the animal is a mammal.
8. The method of Claim 7, wherein the mammal is a human.
9. The method of any preceding claim, wherein the *E. coli* includes one or a plurality of: intestinal pathotypes EAEC, EHEC, enteroinvasive *E. coli* (EIEC), adherent-invasive *E. coli* (AIEC), enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC) enteroaggregative and hemorrhagic *E. coli* (EAHEC) pathotypes, extraintestinal pathotypes comprising avian-pathogenic *E. coli* (APEC), neonatal meningitis *E. coli* (NMEC) and uropathogenic *E. coli* (UPEC).
10. The method of any preceding claim, wherein the isolated protein of SEQ ID NO:1, or one or more immunogenic fragments variants and/or derivatives thereof, or the isolated nucleic acid is administered in combination with one or more additional immunogens that are specific to certain strains/pathotypes of *E. coli*, or are conserved across a number of different *E. coli* strains/pathotypes.
11. The method of Claim 10, wherein the other immunogens are, or include, iron-receptor proteins, fimbrial proteins, flagella, capsular antigens and O antigens.
12. An immunogenic fragment of an isolated protein comprising the amino acid sequence of SEQ ID NO:1 that comprises one or more T and/or B epitopes.
13. An antibody or antibody fragment that binds or is raised against an isolated protein comprising the amino acid sequence of SEQ ID NO:1, or an immunogenic fragment of the isolated protein that preferably comprises one or more T and/or B epitopes.
14. An isolated nucleic acid that comprises a nucleotide sequence that encodes the immunogenic fragment of Claim 12, or a nucleotide sequence complementary

thereto.

15. A genetic construct comprising the isolated nucleic acid of Claim 14, operably linked or connected to one or more regulatory sequences in an expression vector.
16. A host cell transformed or transfected with the genetic construct of Claim 15.
- 5 17. The immunogenic fragment of Claim 12, the isolated nucleic acid of Claim 14, or the genetic construct of Claim 15, for use according to the method of any one of Claims 1-11.
- 10 18. A composition comprising: (a) an isolated protein comprising the amino acid sequence of SEQ ID NO:1; (b) the immunogenic fragment of Claim 12; (c) a variant or derivative of (a) or (b); (d) an antibody or antibody fragment that binds (a), (b) or (c); (e) an isolated nucleic acid encoding (a), (b) or (c); (f) the genetic construct of Claim 15; or (g) the host cell of Claim 16.
- 15 19. The composition of Claim 18, further comprising one or more additional immunogens that are specific to certain strains/pathotypes of *E coli*, or are conserved across a number of different *E coli* strains/pathotypes.
20. The composition of Claim 18 or Claim 19, for use according to the method of any one Claims 1-11.

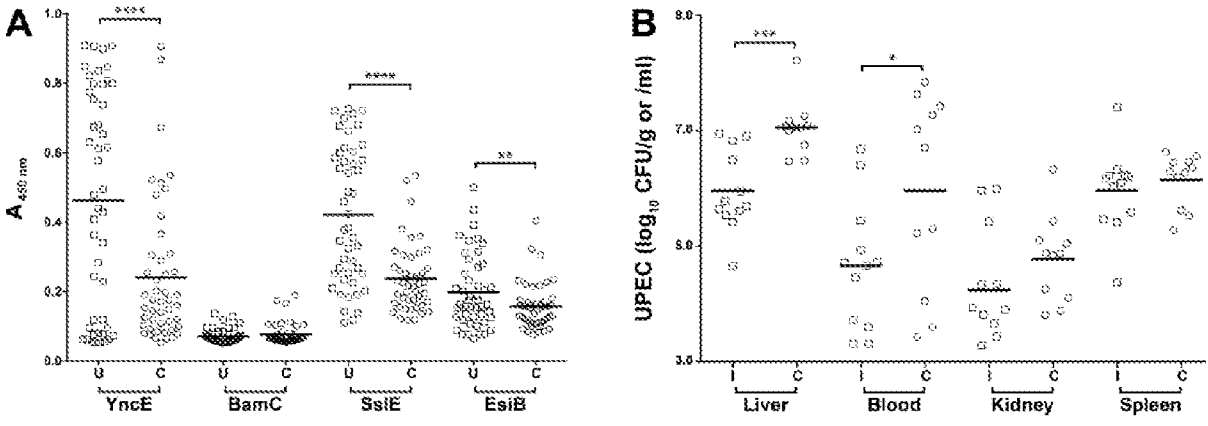


Figure 1

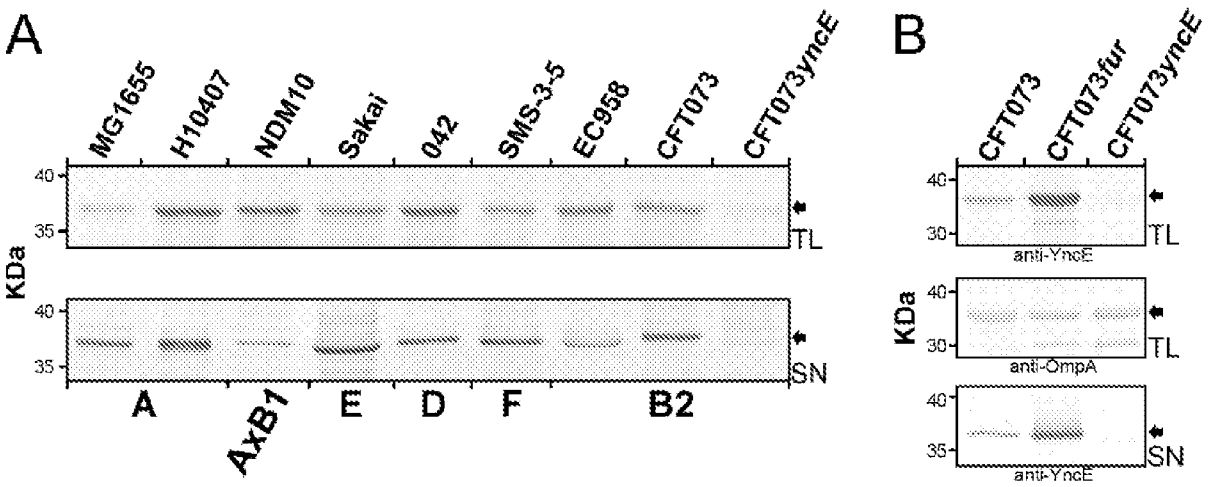


Figure 2

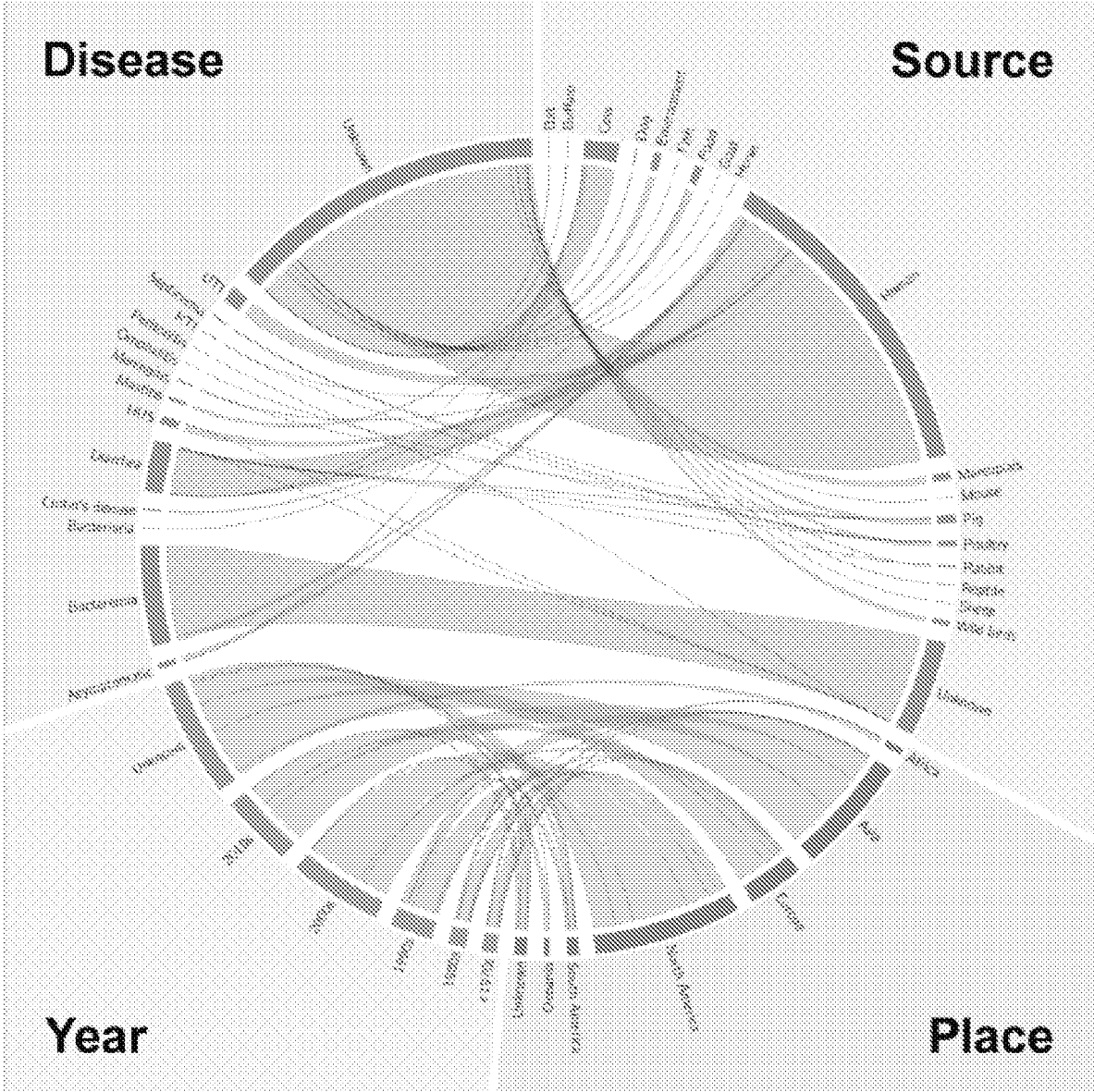


Figure 3

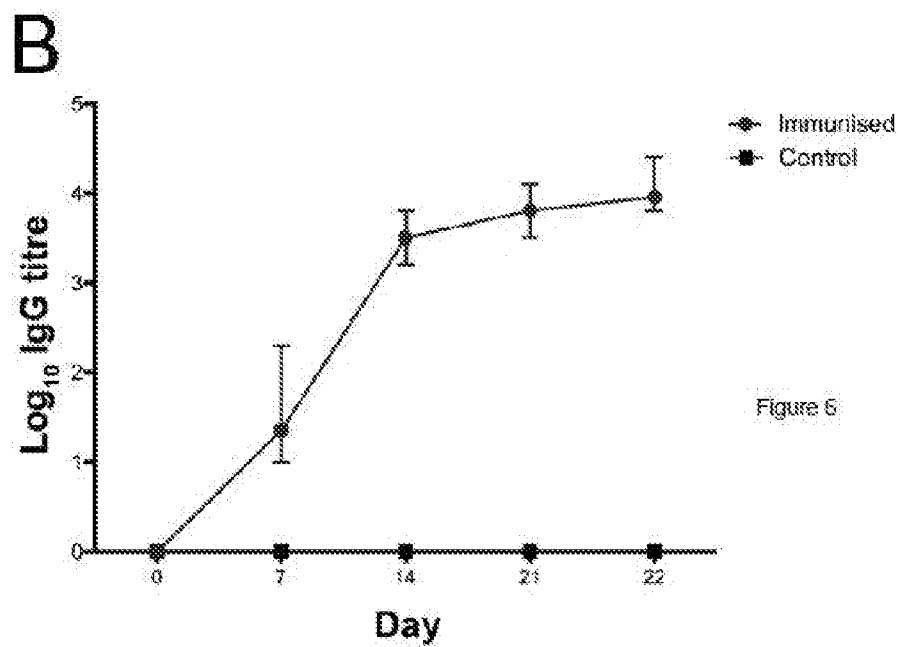
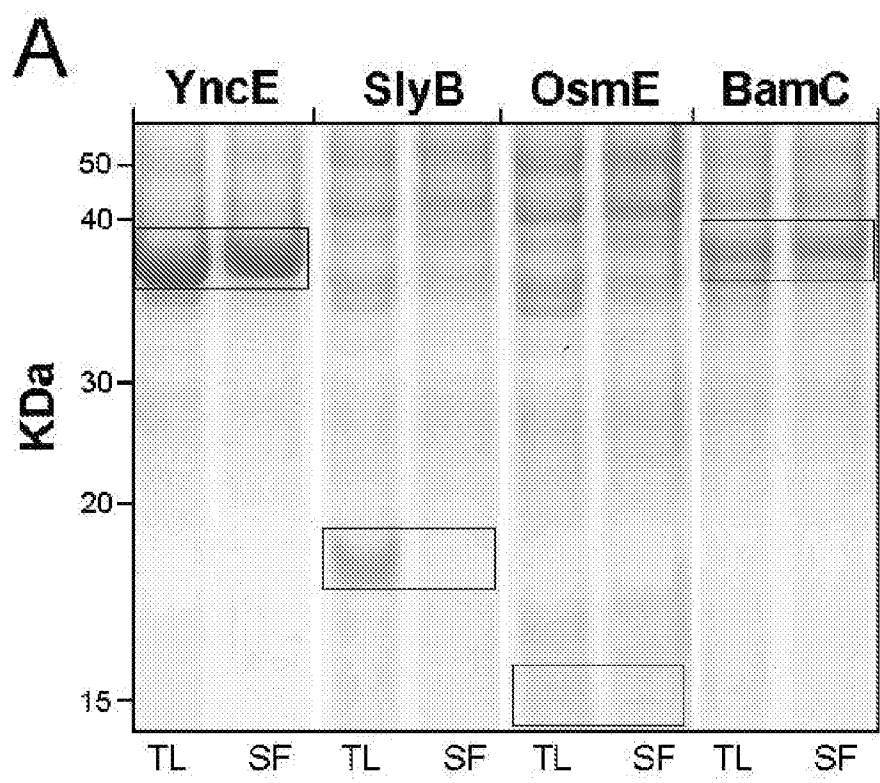


Figure 6

Figure 4

Dataset 8 Primers used in this study		
Construct	Primer	5'-3' sequence*
<i>Protein expression</i>		
pMCSG7-YncE	4636	TACTTCCAATCCAATGCGGCAGAAAGAAATGCTGCGTAAAG
	4637	TTATCCACTTCCAATGTTACAGCGCAATACGAATCACATC
pMCSG7-SlyB	4640	TACTTCCAATCCAATGCGGTTAATAAGACAGCTGTCTAG
	4641	TTATCCACTTCCAATGTTAGCGGCGGGAACGG
pMCSG7-OsmE	4642	TACTTCCAATCCAATGCGACGGCTTATGATCGTACCAA
	4643	TTATCCACTTCCAATGTTACTTCGTAGCCTGCGGATC
pMCSG7-BamE	4650	TACTTCCAATCCAATGCGAGTTCTGALTCACTGCTATAAGC
	4651	TTATCCACTTCCAATGTTACTTGCTAAACGCAGCCTG
<i>Knockout generation</i>		
CFT073yncE	5600	GCGAAGCATTGTGAAGCAGCAAAAATATCGGTTTCATCAAAGGGAGTGGTCGTGTAGGETGGAGCTGETTC
	5601	GCGTGATAAAAGTCAGGCAATTCCTTCCGTTTTACCCCTCGCLATATCTACATATGAATATCCTCTTAG

Figure 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU2017/051164

A. CLASSIFICATION OF SUBJECT MATTER

A61K 39/108 (2006.01) C07K 14/245 (2006.01) C12N 15/31 (2006.01) C12Q 1/10 (2006.01)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PATENW, MEDLINE, CAPLUS, BIOSIS, EMBASE, GENOME QUEST, PATENTSCOPE, AUSPAT, GOOGLE, GOOGLE ADVANCED PATENT SEARCH, GOOGLE SCHOLAR, ESPACENET, PUBMED (YncE, ECK 1446, JW 1447, SEQ ID NO: 1, The University of Queensland; SCHEMBRI M; MORIEL D)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	Documents are listed in the continuation of Box C	

☒ Further documents are listed in the continuation of Box C ☒ See patent family annex

* "A"	Special categories of cited documents: document defining the general state of the art which is not considered to be of particular relevance	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E"	earlier application or patent but published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means	"&"	document member of the same patent family
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 5 January 2018	Date of mailing of the international search report 05 January 2018
Name and mailing address of the ISA/AU AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA Email address: pct@ipaustalia.gov.au	Authorised officer Anita Premkumar AUSTRALIAN PATENT OFFICE (ISO 9001 Quality Certified Service) Telephone No. +61262832013

INTERNATIONAL SEARCH REPORT		International application No.
C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		PCT/AU2017/051164
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2004/0029129 A1 (WANG, L et al.) 12 February 2004 Abstract, [1138]-[1139], [0150]-[0155], [0250]-[0252], [1458]; claims 25, 32, 39, 58-68, 90, 96, 104; Example 12 and 86, SEQ ID NOs: 6895 and 43079	1-20
X	WO 2004/018638 A2 (REGENTS OF THE UNIVERSITY OF MINNESOTA) 04 March 2004 Abstract, pages 4-6, page 23 line 15 - page 28, page 30 lines 1-5, claims 1, 7, 10, Fig 4, SEQ ID NOs: 27 and 91	1-20
X	BABA-DIKWA, A et al. "Overproduction, purification and preliminary X-ray diffraction analysis of YncE, an iron-regulated Sec-dependent periplasmic protein from <i>Escherichia coli</i> ." Acta Cryst. (2008). F64, pages 966-969 Abstract, page 966 para 1, page 967 col 1 para 1	1-20
X	WURPEL, D J et al. "Comparative proteomics of uropathogenic <i>Escherichia coli</i> during growth in human urine identify UCA-like (UCL) fimbriae as an adherence factor involved in biofilm formation and binding to uroepithelial cells." Journal of Proteomics (Available online 3 November 2015), VOL 131, pages 177-189. Tables 1 and 2, Fig 2	1-20
X	GenBank Accession No. ALI45170; 16 December 2015 Sequence shares 100% identity with SEQ ID NO: 1	1-20
X	GenBank Accession No. ALB31585; 31 August 2015 Sequence shares 100% identity with SEQ ID NO: 1	1-20

Form PCT/ISA/210 (fifth sheet) (July 2009)

INTERNATIONAL SEARCH REPORT Information on patent family members		International application No. PCT/AU2017/051164	
This Annex lists known patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.			
Patent Document/s Cited in Search Report		Patent Family Member/s	
Publication Number	Publication Date	Publication Number	Publication Date
US 2004/0029129 A1	12 February 2004	US 2004029129 A1	12 Feb 2004
		AU 4934501 A	03 Oct 2001
		AU 2002241706 A1	08 Jul 2002
		AU 2002306849 A1	08 Oct 2002
		CA 2404260 A1	27 Sep 2001
		EP 1268774 A2	02 Jan 2003
		JP 2004516805 A	10 Jun 2004
		JP 4852211 B2	11 Jan 2012
		KR 20020097200 A	31 Dec 2002
		US 2002061569 A1	23 May 2002
		US 2003027286 A1	06 Feb 2003
		WO 0170955 A2	27 Sep 2001
		WO 02051982 A2	04 Jul 2002
		WO 02077183 A2	03 Oct 2002
WO 2004/018638 A2	04 March 2004	WO 2004018638 A2	04 Mar 2004
		AU 2003260043 A1	11 Mar 2004
End of Annex			
Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001. Form PCT/ISA/210 (Family Annex)(July 2009)			