

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property Organization

International Bureau

(43) International Publication Date
24 September 2020 (24.09.2020)



(10) International Publication Number

WO 2020/191088 A9

(51) International Patent Classification:
CI2N 9/10 (2006.01) *CI2P 21/00* (2006.01)

(21) International Application Number:
PCT/US2020/023415

(22) International Filing Date:
18 March 2020 (18.03.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/819,762 18 March 2019 (18.03.2019) US

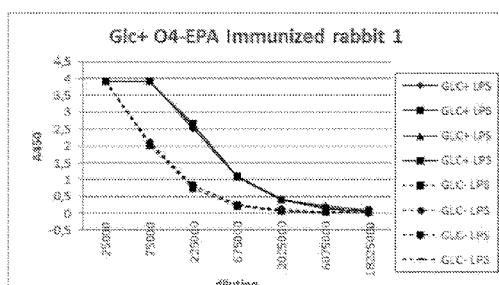
(71) Applicants: JANSSEN PHARMACEUTICALS, INC. [US/US]; 1125 Trenton-Harbourton Road, Titusville, NJ 08560 (US). GLAXOSMITHKLINE BIOLOGICALS S.A. [BE/BE]; Rue de l'Institut 89, B-1330 Rixensart (BE).

(72) Inventors: GEURTSEN, Jeroen; c/o Janssen Vaccines & Prevention B.V., IP Department, Archimedesweg 4-6, 2333 CN Leiden (NL). BURGHOUT, Pieter, Jan; c/o Janssen Vaccines & Prevention B.V., IP Department, Archimedesweg 4-6, 2333 CN Leiden (NL). WEERDEN-

BURG, Eveline, Marleen; c/o Janssen Vaccines & Prevention B.V., IP Department, Archimedesweg 4-6, 2333 CN Leiden (NL). POOLMAN, Jan, Theunis; c/o Janssen Vaccines & Prevention B.V., IP Department, Archimedesweg 4-6, 2333 CN Leiden (NL). FAE, Kellen, Crishina; c/o Janssen Vaccines & Prevention B.V., IP Department, Archimedesweg 4-6, 2333 CN Leiden (NL). IBARRA YON, Patricia; c/o Janssen Vaccines & Prevention B.V., IP Department, Archimedesweg 4-6, 2333 CN Leiden (NL). ABBANAT, Darren, Robert; c/o Janssen Vaccines & Prevention B.V., IP Department, Archimedesweg 4-6, 2333 CN Leiden (NL). KEMMLER, Stefan, Jochen; Grabenstrasse 3, 8952 Schlieren (CH). KOWARIK, Michael, Thomas; Grabenstrasse 3, 8952 Schlieren (CH). MALLY, Manuela; Grabenstrasse 3, 8952 Schlieren (CH). GAMBILLARA, FONCK, Veronica; Grabenstrasse 3, 8952 Schlieren (CH). BRAUN, Martin, Edward; Grabenstrasse 3, 8952 Schlieren (CH). CARRANZA SANDMEIER, Maria, Paula; Grabenstrasse 3, 8952 Schlieren (CH).

(74) Agent: HSING, Weihong et al.; ICE MILLER LLP, 1735 Market Street, Suite 3450, Philadelphia, PA 19103-7509 (US).

(54) Title: METHODS OF PRODUCING BIOCONJUGATES OF *E. COLI* O-ANTIGEN POLYSACCHARIDES, COMPOSITIONS THEREOF, AND METHODS OF USE THEREOF



(57) Abstract: Methods of producing bioconjugates of O-antigen polysaccharides covalently linked to a carrier protein using recombinant host cells are provided. The recombinant host cells used in the methods described herein encode a particular oligosaccharyl transferase enzyme depending on the O-antigen polysaccharide bioconjugate to be produced. The oligosaccharyl transferase enzymes can be PglB oligosaccharyl transferase or variants thereof. Also provided are compositions containing the bioconjugates, and methods of using the bioconjugates and compositions described herein to vaccinate a subject against extra-intestinal pathogenic *E. coli*. (ExPEC).

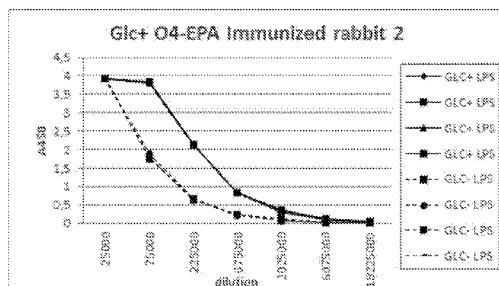


Fig. 1



-
- (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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

- *with international search report (Art. 21(3))*
- *with sequence listing part of description (Rule 5.2(a))*

(48) Date of publication of this corrected version:

18 March 2021 (18.03.2021)

(15) Information about Correction:

see Notice of 18 March 2021 (18.03.2021)

Previous Correction:

see Notice of 05 November 2020 (05.11.2020)

TITLE OF THE INVENTION

Methods of Producing Bioconjugates of *E. coli* O-Antigen Polysaccharides, Compositions Thereof, and Methods of Use Thereof

CROSS REFERENCE TO RELATED APPLICATION

[001] This application claims priority to U.S. Provisional Application No. 62/819,762 filed on March 18, 2019, the disclosure of which is incorporated herein by reference in its entirety.

REFERENCE TO SEQUENCE LISTING SUBMITTED ELECTRONICALLY

[002] This application contains a sequence listing, which is submitted electronically via EFS-Web as an ASCII formatted sequence listing with a file name “004852_11612 Sequence_Listing”, creation date of March 11, 2020, and having a size of 199 KB. The sequence listing submitted via EFS-Web is part of the specification and is herein incorporated by reference in its entirety.

BACKGROUND OF THE INVENTION

[003] Extraintestinal pathogenic *Escherichia coli* (ExPEC) strains are normally harmless inhabitants of the human gastrointestinal tract, alongside commensal *E. coli* strains. ExPEC isolates cannot readily be distinguished from commensal isolates by serotype, although many clonal lineages are dominated by ExPEC, as defined by O-antigen, capsule and flagellar antigen serotypes (abbreviated as O:K:H, for example O25:K1:H4). In contrast to commensal *E. coli*, ExPEC strains express a broad array of virulence factors enabling them to colonize the gastrointestinal tract, as well as to cause a wide range of extraintestinal infections, which are associated with a significant healthcare cost burden due to hospitalization and death. Neonates, the elderly, and immunocompromised patients are particularly susceptible to ExPEC infection, including invasive ExPEC disease (IED).

[004] ExPEC strains are the most common cause of urinary tract infections (UTI) and important contributors to surgical site infections and neonatal meningitis. The strains are also associated with abdominal and pelvic infections and nosocomial pneumonia, and are occasionally involved in other extraintestinal infections, such as osteomyelitis, cellulitis, and wound infections. All these primary sites of infection can result in ExPEC bacteremia. ExPEC is the most common cause of community-onset bacteremia and a major causative pathogen in nosocomial bacteremia and is found in about 17% to 37% of clinically significant blood isolates. Patients with an ExPEC-positive blood culture typically suffer sepsis syndrome, severe sepsis, or septic shock. Increasing resistance of ExPEC against first-line antibiotics including the cephalosporins, fluoroquinolones, and trimethoprim/sulfamethoxazole has been observed. The emergence and rapid global dissemination of ExPEC sequence type 131 (ST131) is considered a main driver of increased drug resistance, including multi-drug resistance. This clone is found in 12.5% to 30% of all ExPEC clinical isolates, exhibits mostly serotype O25b:H4, and shows high levels of resistance to fluoroquinolones, which is often accompanied by trimethoprim/sulfamethoxazole resistance and extended-spectrum beta-lactamases conferring resistance to cephalosporins.

[005] The O-antigen comprises the immunodominant component of the cell wall lipopolysaccharide (LPS) in Gram-negative bacteria, including *E. coli*. There are currently >180 serologically unique *E. coli* O-antigens identified, with the vast majority of ExPEC isolates classified within less than 20 O-antigen serotypes. Full-length *E. coli* O-antigens are typically comprised of about 10 to 25 repeating sugar units attached to the highly conserved LPS core structure, with each component synthesized separately by enzymes encoded predominantly in the *rfb* and *rfa* gene clusters, respectively. Following polymerization of the O-antigen, the O-antigen polysaccharide backbone may be modified, typically through the addition of acetyl or glucose residues. These modifications effectively increase serotype diversity by creating antigenically distinct serotypes that share a common polysaccharide backbone, but differ in side branches. Genes encoding O-antigen modifying enzymes typically reside outside of the *rfb* cluster on the chromosome, and in some cases, these genes are found within lysogenic bacteriophages.

[006] ExPEC isolates belonging to the O4 serogroup have been commonly identified in contemporary surveillance studies of U.S. and EU blood isolates. The structure of the O4

polysaccharide was determined as -->2) α -L-Rha (1->6) α -D-Glc (1->3) α -L-FucNAc (1->3) β -D-GlcNAc (1-> from an *E. coli* O4:K52 strain (Jann et al., *Carbohydr. Res.* (1993) v. 248, pp.241-250). A distinct form of the O4 polysaccharide structure was determined for O4:K3, O4:K6 and O4:K12 strains, in which the structure above was modified by the addition of an α -D-Glc (1->3) linked to the rhamnose residue of the polysaccharide (Jann et al., 1993, *supra*), this form of the polysaccharide referred to herein below as ‘glucosylated O4’. The enzymes responsible for the O-antigen modification within *E. coli* O4 strains were not identified.

[007] Efforts toward the development of a vaccine to prevent ExPEC infections have focused on O-antigen polysaccharide conjugates. A 12-valent O-antigen conjugate vaccine was synthesized through extraction and purification of O-antigen polysaccharide and chemical conjugation to detoxified *Pseudomonas aeruginosa* exotoxin A and tested for safety and immunogenicity in a Phase 1 clinical study (Cross et al., *J. Infect. Dis.* (1994) v.170, pp.834-40). This candidate vaccine was never licensed for clinical use. A bioconjugation system in *E. coli* has been developed recently, in which the polysaccharide antigen and the carrier protein are both synthesized *in vivo* and subsequently conjugated *in vivo* through the activities of the oligosaccharyl transferase PglB, a *Campylobacter jejuni* enzyme, expressed in *E. coli* (Wacker et al., *Proc. Nat. Acad. Sci.* (2006) v. 103, pp. 7088-93). This N-linked protein glycosylation system is capable of the transfer of diverse polysaccharides to a carrier protein, allowing for straightforward methods to purify the conjugate.

[008] Bioconjugation has been used successfully to produce conjugate polysaccharide for an *E. coli* four-valent O-antigen candidate vaccine (Poolman and Wacker, *J. Infect. Dis.* (2016) v.213(1), pp. 6-13). However, the development of a successful ExPEC vaccine requires coverage of predominant serotypes, and the presence of further O-antigen modifications in subsets of ExPEC isolates presents a further challenge in covering isolates displaying unmodified and modified LPS. Moreover, efficiency of production of the multiple components for more complex vaccine compositions covering multiple serotypes becomes increasingly important, and hence there remains a need for improvements in production of individual bioconjugates of specific O-antigens.

BRIEF SUMMARY OF THE INVENTION

[009] In view of increasing antibiotic resistance among ExPEC isolates and the presence of further O-antigen modifications among predominant O-serotypes, there is a need for improved prophylactic and therapeutic treatments for these infections. The invention satisfies this need by defining the genetic composition of contemporary clinical isolates, including identifying the genes encoding O-antigen modifying enzymes, thus allowing for the engineering of recombinant host cells capable of synthesizing bioconjugates of the O-antigens including bioconjugates comprising selected O-antigen modifications. In addition, in one aspect of the invention, host cells and methods for improved production of bioconjugates of specific O-antigens by using variants of oligosaccharyltransferase (OST) are provided, based on advantages of use of certain OST variants for bioconjugates of certain *E. coli* O-antigens in an unpredictable serotype-dependent manner. Use of such OST variants may in certain cases also affect the glycosylation pattern of the bioconjugate, e.g. by increasing the relative number of glycans coupled to the carrier protein as compared to bioconjugates produced using wild-type or other variants of the OST, and hence novel bioconjugates produced by such methods are also provided as an aspect of the invention.

[0010] In one aspect, provided is a method of preparing a bioconjugate of an *E. coli* O_x antigen polysaccharide covalently linked to a carrier protein, the method comprising:

(i) providing a recombinant host cell comprising:

- a. a nucleotide sequence of an *rfb* gene cluster for the O_x-antigen polysaccharide;
- b. a nucleotide sequence encoding the carrier protein comprising at least one glycosylation site comprising a glycosylation consensus sequence having SEQ ID NO: 1, preferably having SEQ ID NO: 2; and
- c. a nucleotide sequence encoding an oligosaccharyl transferase PglB; and

(ii) culturing the recombinant host cell under conditions for production of the bioconjugate,

wherein:

when the O_x-antigen is O1A antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V;

when the O_x-antigen is glucosylated O4 antigen polysaccharide, the PglB_y comprises the amino acid mutation N311V or the amino acid mutations Y77H and N311V, and the recombinant host cell further comprises a sequence encoding a glucosyltransferase GtrS having at least 80% identity to SEQ ID NO: 4 and being capable of modifying an *E. coli* O4 antigen polysaccharide by addition of glucose to produce the *E. coli* glucosylated O4 antigen polysaccharide, and nucleotide sequences encoding a translocase GtrA and a glycosyltransferase GtrB having at least 80% sequence identity to SEQ ID NOs: 7 and 8 respectively, wherein the translocase is capable of translocating bactoprenol linked glucose and the glycosyltransferase is capable of glucosylating bactoprenol;

when the O_x-antigen is O6A antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V;

when the O_x-antigen is O8 antigen polysaccharide, the PglB_y comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483 and 669;

when the O_x-antigen is O15 antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V;

when the O_x-antigen is O16 antigen polysaccharide, the PglB_y comprises the amino acid mutations of Y77H, S80R, Q287P, K289R, and N311V;

when the O_x-antigen is O18A antigen polysaccharide, the PglB_y comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483 and 669; and

when the O_x-antigen is O75 antigen polysaccharide, the PglB_y comprises the amino acid mutation of N311V,

wherein in each case the amino acid mutations are relative to the wild-type PglB having the amino acid sequence of SEQ ID NO: 6,

wherein the O1A, glucosylated O4, O6A, O8, O15, O16, O18A, and O75 antigen polysaccharides have the structures of Formulas (O1A), (O4-Glc+), (O6A), (O8), (O15), (O16), (O18A), and (O75), respectively, as shown in Table 1, and each n is independently an integer of 1 to 100, preferably 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20.

[0011] In one embodiment, the O_x-antigen is O1A antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[0012] In one embodiment, the O_x-antigen is glucosylated O4 antigen polysaccharide, and the PglB_y comprises the amino acid mutation N311V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6. In one embodiment, the O_x-antigen is glucosylated O4 antigen polysaccharide, and the PglB_y comprises the amino acid mutations Y77H and N311V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6. In embodiments wherein the O_x-antigen is glucosylated O4 antigen polysaccharide, the recombinant host cell preferably further comprises a sequence encoding a GtrS having at least 80% identity to SEQ ID NO: 4, and nucleotide sequences encoding a GtrA and a GtrB having at least 80% sequence identity to SEQ ID NOs: 7 and 8 respectively.

[0013] In one embodiment, the O_x-antigen is O6A antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[0014] In one embodiment, the O_x-antigen is O8 antigen polysaccharide, the PglB_y comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483 and 669 relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[0015] In one embodiment, the O_x-antigen is O15 antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[0016] In one embodiment, the O_x-antigen is O16 antigen polysaccharide, the PglB_y comprises the amino acid mutations of Y77H, S80R, Q287P, K289R, and N311V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[0017] In one embodiment, the O_x-antigen is O18A antigen polysaccharide, the PglB_y comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483 and 669 relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6, and preferably comprises the amino acid sequence of SEQ ID NO: 6.

[0018] In one embodiment, the O_x-antigen is O75 antigen polysaccharide, the PgIB_y comprises the amino acid mutation of N311V relative to wild-type PgIB having the amino acid sequence of SEQ ID NO: 6.

[0019] In a particular aspect, provided is a method of preparing a bioconjugate of an *E. coli* O_x-antigen polysaccharide covalently linked to a carrier protein, the method comprising:

(i) providing a recombinant host cell comprising:

- a. a nucleotide sequence of an *rfb* gene cluster for the O_x-antigen polysaccharide;
- b. a nucleotide sequence encoding the carrier protein comprising at least one glycosylation site comprising a glycosylation consensus sequence having SEQ ID NO: 1, preferably having SEQ ID NO: 2; and
- c. a nucleotide sequence encoding an oligosaccharyl transferase PgIB_y; and

(ii) culturing the recombinant host cell under conditions for production of the bioconjugate,

wherein the PgIB_y comprises the amino acid mutation N311V relative to the wild-type PgIB having the amino acid sequence of SEQ ID NO: 6,

wherein the O_x- antigen is O1A antigen polysaccharide, glucosylated O4 antigen polysaccharide, O6A antigen polysaccharide, O15 antigen polysaccharide, O16 antigen polysaccharide, or O75 antigen polysaccharide, and when the O_x-antigen is glucosylated O4 antigen polysaccharide, the recombinant host cell further comprises a sequence encoding a glucosyltransferase GtrS having at least 80% identity to SEQ ID NO: 4 and being capable of modifying an *E. coli* O4 antigen polysaccharide by addition of glucose to produce the *E. coli* glucosylated O4 antigen polysaccharide, and nucleotide sequences encoding a translocase GtrA and a glycosyltransferase GtrB having at least 80% sequence identity to SEQ ID NOS: 7 and 8, respectively, wherein the translocase is capable of translocating bactoprenol linked glucose and the glycosyltransferase is capable of glucosylating bactoprenol, and

wherein the O1A, glucosylated O4, O6A, O15, O16, and O75 antigen polysaccharides have the structures of Formulas (O1A), (O4-Glc+), (O6A), (O15), (O16), and (O75), respectively, as shown in Table 1, and each n is independently an integer of 1 to 100, preferably 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20.

[0020] In certain embodiments, the method further comprises isolating the bioconjugate from the recombinant host cell.

[0021] In certain embodiments, the carrier protein protein is selected from the group consisting of detoxified Exotoxin A of *P. aeruginosa* (EPA), *E. coli* flagellin (FliC), CRM197, maltose binding protein (MBP), Diphtheria toxoid, Tetanus toxoid, detoxified hemolysin A of *S. aureus*, clumping factor A, clumping factor B, *E. coli* heat labile enterotoxin, detoxified variants of *E. coli* heat labile enterotoxin, Cholera toxin B subunit (CTB), cholera toxin, detoxified variants of cholera toxin, *E. coli* Sat protein, the passenger domain of *E. coli* Sat protein, *Streptococcus pneumoniae* Pneumolysin, Keyhole limpet hemocyanin (KLH), *P. aeruginosa* PcrV, outer membrane protein of *Neisseria meningitidis* (OMPC), and protein D from non-typeable *Haemophilus influenzae*.

[0022] In certain embodiments, the carrier protein is detoxified exotoxin A of *Pseudomonas aeruginosa* (EPA). Preferably, the EPA carrier protein comprises 1-10, preferably 2-4, more preferably 4, glycosylation sites. In certain embodiments, each glycosylation site comprises a glycosylation consensus sequence having SEQ ID NO: 2. In a particular embodiment, the EPA carrier protein comprises SEQ ID NO: 3.

[0023] In certain embodiments, the recombinant host cell is an *E. coli* cell, e.g., an *E. coli* K-12 strain, such as strain W3110.

[0024] In another aspect, provided is a bioconjugate produced by a method of preparing a bioconjugate of an O_x antigen polysaccharide covalently linked to a carrier protein as described herein.

[0025] In another aspect, provided is a composition comprising such a bioconjugate. In some embodiments, a composition comprises at least 2, preferably at least 3, more preferably at least 5, still more preferably at least 7 of such bioconjugates.

[0026] In certain embodiments, a composition according to the invention comprises a bioconjugate of *E. coli* glucosylated O4 antigen polysaccharide covalently linked to a carrier

protein, wherein the glucosylated O4 antigen polysaccharide has the structure of Formula (O4-Glc+) as shown in Table 1, and n is an integer of 1 to 100, preferably 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20. In certain embodiments, a composition according to the invention further comprises at least a bioconjugate of *E. coli* O25B antigen polysaccharide covalently linked to a carrier protein, wherein the O25B antigen polysaccharide has the structure of Formula (O25B) as shown in Table 1, and n is an integer of 1 to 100, preferably 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20. In certain embodiments, a composition according to the invention further comprises at least a bioconjugate of *E. coli* O2 antigen polysaccharide covalently linked to a carrier protein, wherein the O2 antigen polysaccharide has the structure of Formula (O2) as shown in Table 1, and n is an integer of 1 to 100, preferably 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20.

[0027] In certain embodiments, a composition of the invention comprises: (i) bioconjugate of *E. coli* O1A antigen polysaccharide covalently coupled to a carrier protein, (ii) bioconjugate of *E. coli* O2 antigen polysaccharide covalently coupled to a carrier protein, (iii) bioconjugate of *E. coli* glucosylated O4 antigen polysaccharide covalently coupled to a carrier protein, (iv) bioconjugate of *E. coli* O6A antigen polysaccharide covalently coupled to a carrier protein, (v) bioconjugate of *E. coli* O8 antigen polysaccharide covalently coupled to a carrier protein, (vi) bioconjugate of *E. coli* O15 antigen polysaccharide covalently coupled to a carrier protein, (vii) bioconjugate of *E. coli* O16 antigen polysaccharide covalently coupled to a carrier protein, (viii) bioconjugate of *E. coli* O25B antigen polysaccharide covalently coupled to a carrier protein, and (ix) bioconjugate of *E. coli* O75 antigen polysaccharide covalently coupled to a carrier protein, wherein the O1A, O2, glucosylated O4, O6A, O8, O15, O16, O25B, and O75 antigen polysaccharides have the structures of Formulas (O1A), (O2), (O4-Glc+), (O6A), (O8), (O15), (O16), (O25B), and (O75), respectively, as shown in Table 1, and each n is independently an integer of 1 to 100, preferably 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20. In certain embodiments, such a composition further comprises: (x) bioconjugate of *E. coli* O18A antigen polysaccharide covalently coupled to a carrier protein, wherein the O18A antigen polysaccharide has the structure of Formula (O18A) as shown in Table 1, and n is an integer of 1 to 100, preferably 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20. In certain embodiments, a composition of the invention is an immunogenic composition.

[0028] In other aspects, provided is a method of vaccination a subject against extra-intestinal pathogenic *E. coli* (ExPEC), comprising administering to the subject such a bioconjugate or composition as described herein. In yet other aspects, provided is such bioconjugate or composition as described herein for use in vaccination against extra-intestinal pathogenic *E. coli* (ExPEC).

[0029] In other aspects, provided are recombinant host cells for preparing a bioconjugate of an *E. coli* O_x antigen polysaccharide covalently linked to a carrier protein, the recombinant host cell comprising:

- (a) a nucleotide sequence of an *rfb* gene cluster for the O_x-antigen polysaccharide;
- (b) a nucleotide sequence encoding the carrier protein comprising at least one glycosylation site comprising a glycosylation consensus sequence having SEQ ID NO: 1, preferably having SEQ ID NO: 2; and
- (c) a nucleotide sequence encoding an oligosaccharyl transferase PglB_y,

wherein:

when the O_x-antigen is O1A antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V;

when the O_x-antigen is glucosylated O4 antigen polysaccharide, the PglB_y comprises the amino acid mutation N311V or the amino acid mutations Y77H and N311V, and the recombinant host cell further comprises a sequence encoding a glucosyltransferase GtrS having at least 80% identity to SEQ ID NO: 4 and being capable of modifying an *E. coli* O4 antigen polysaccharide by addition of glucose to produce the *E. coli* glucosylated O4 antigen polysaccharide, and nucleotide sequences encoding a translocase GtrA and a glycosyltransferase GtrB having at least 80% sequence identity to SEQ ID NOs: 7 and 8 respectively, wherein the translocase is capable of translocating bactoprenol linked glucose and the glycosyltransferase is capable of glucosylating bactoprenol;

when the O_x-antigen is O6A antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V;

when the O_x-antigen is O6A antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V;

when the O_x-antigen is O15 antigen polysaccharide, the PglB_y comprises the amino acid

mutations of N311V, K482R, D483H, and A669V;

when the O_x-antigen is O16 antigen polysaccharide, the PglB_y comprises the amino acid mutations of Y77H, S80R, Q287P, K289R, and N311V;

when the O_x-antigen is O18A antigen polysaccharide, the PglB_y comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483 and 669; and

when the O_x-antigen is O75 antigen polysaccharide, the PglB_y comprises the amino acid mutation of N311V,

wherein in each case the amino acid mutations are relative to the wild-type PglB having the amino acid sequence of SEQ ID NO: 6, and

wherein the O1A, glucosylated O4, O6A, O8, O15, O16, O18A, and O75 antigen polysaccharides have the structures of Formulas (O1A), (O4-Glc+), (O6A), (O8), (O15), (O16), (O18A), and (O75), respectively, as shown in Table 1, and each n is independently an integer of 1 to 100, preferably 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20.

[0030] In certain embodiments, such host cells are provided wherein the O_x-antigen is O1A antigen polysaccharide, and the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[0031] In certain embodiments, recombinant host cells of the invention are provided wherein the O_x-antigen is glucosylated O4 antigen polysaccharide, and the PglB_y comprises the amino acid mutation N311V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6. In certain embodiments, recombinant host cells of the invention are provided wherein the O_x-antigen is glucosylated O4 antigen polysaccharide, and the PglB_y comprises the amino acid mutations Y77H and N311V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6. In certain embodiments wherein the O_x-antigen is glucosylated O4 antigen polysaccharide, the recombinant host cell further comprises a sequence encoding a GtrS having the amino acid sequence of SEQ ID NO: 4, and nucleotide sequences encoding a GtrA and a GtrB having the amino acid sequences of SEQ ID NOs: 7 and 8, respectively.

[0032] In certain embodiments, recombinant host cells of the invention are provided wherein the O_x-antigen is O6A antigen polysaccharide, and the PglB_y comprises the amino acid mutations

of N311V, K482R, D483H, and A669V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[0033] In certain embodiments, recombinant host cells of the invention are provided wherein the O_x-antigen is O8 antigen polysaccharide, the PglB_y comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483 and 669 relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[0034] In certain embodiments, recombinant host cells of the invention are provided wherein the O_x-antigen is O15 antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[0035] In certain embodiments, recombinant host cells of the invention are provided wherein the O_x-antigen is O16 antigen polysaccharide, the PglB_y comprises the amino acid mutations of Y77H, S80R, Q287P, K289R, and N311V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[0036] In certain embodiments, recombinant host cells of the invention are provided wherein the O_x-antigen is O18A antigen polysaccharide, the PglB_y comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483 and 669 relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[0037] In certain embodiments, recombinant host cells of the invention are provided wherein the O_x-antigen is O75 antigen polysaccharide, the PglB_y comprises the amino acid mutation of N311V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[0038] In certain embodiments, recombinant host cells of the invention are provided wherein the carrier protein is selected from the group consisting of detoxified Exotoxin A of *P. aeruginosa* (EPA), *E. coli* flagellin (FliC), CRM197, maltose binding protein (MBP), Diphtheria toxoid, Tetanus toxoid, detoxified hemolysin A of *S. aureus*, clumping factor A, clumping factor B, *E. coli* heat labile enterotoxin, detoxified variants of *E. coli* heat labile enterotoxin, Cholera toxin B subunit (CTB), cholera toxin, detoxified variants of cholera toxin, *E. coli* Sat protein, the passenger domain of *E. coli* Sat protein, *Streptococcus pneumoniae* Pneumolysin, Keyhole limpet hemocyanin (KLH), *P. aeruginosa* PcrV, outer membrane protein of *Neisseria meningitidis* (OMPC), and protein D from non-typeable *Haemophilus influenzae*.

[0039] In certain embodiments, recombinant host cells of the invention are provided wherein the carrier protein is detoxified exotoxin A of *Pseudomonas aeruginosa* (EPA). In certain embodiments thereof, the EPA carrier protein comprises 1-10, preferably 2-4, more preferably 4, of the glycosylation sites. In certain embodiments, each glycosylation site comprises a glycosylation consensus sequence having SEQ ID NO: 2. In certain embodiments, the EPA carrier protein comprises SEQ ID NO: 3.

[0040] In certain embodiments, recombinant host cells of the invention are provided wherein the recombinant host cell is an *E. coli* cell, e.g. an *E. coli* K-12 strain, such as strain W3110.

[0041] In certain embodiments for the host cells and methods for preparing a bioconjugate of an *E. coli* glucosylated O4 antigen polysaccharide covalently linked to a carrier protein according to the invention, the *rfb* gene cluster for the *E. coli* O4 antigen polysaccharide comprises a sequence that encodes the enzymes that create the *E. coli* O4 antigen polysaccharide (Formula (O4-Glc-) in Table 1) and is at least 80%, e.g. at least 90%, e.g. at least 95%, e.g. at least 98% identical to SEQ ID NO: 9. In certain embodiments the *rfb* gene cluster comprises SEQ ID NO: 9.

[0042] In certain embodiments for the host cells and methods for preparing a bioconjugate of an *E. coli* glucosylated O4 antigen polysaccharide covalently linked to a carrier protein according to the invention, the glucosyl transferase that is capable of modifying the *E. coli* O4 antigen polysaccharide to produce the *E. coli* glucosylated O4 antigen polysaccharide has an amino acid sequence that has at least 90%, preferably at least 95%, preferably at least 98% sequence identity to SEQ ID NO: 4. In certain embodiments, the glucosyl transferase comprises SEQ ID NO: 4.

[0043] In certain embodiments for the host cells and methods for preparing a bioconjugate of an *E. coli* glucosylated O4 antigen polysaccharide covalently linked to a carrier protein according to the invention, the translocase is capable of translocating bactoprenol-linked glucose and has at least 90%, preferably at least 95%, preferably at least 98% sequence identity to SEQ ID NO: 7. In certain embodiments, the translocase comprises SEQ ID NO: 7.

[0044] In certain embodiments for the host cells and methods for preparing a bioconjugate of an *E. coli* glucosylated O4 antigen polysaccharide covalently linked to a carrier protein according to the invention, the glycosyltransferase is capable of glucosylating bactoprenol and

has at least 90%, preferably at least 95%, preferably at least 98% sequence identity to SEQ ID NO: 8. In certain embodiments, the glycosyltransferase comprises SEQ ID NO: 8.

BRIEF DESCRIPTION OF THE FIGURES

[0045] The foregoing summary, as well as the following detailed description of the invention, will be better understood when read in conjunction with the appended drawings. It should be understood that the invention is not limited to the precise embodiments shown in the drawings.

[0046] In the drawings:

[0047] **FIG. 1** shows ELISA IgG titers against unmodified (GLC-) or glucose-modified (GLC+) O4 LPS in sera from two rabbits immunized with Glc-modified O4 polysaccharide bioconjugate as described in Example 4; ELISA titers were determined in quadruplicate;

[0048] **FIG. 2** shows IgG titers in whole cell ELISAs with pooled sera from rabbits immunized with a Glc-modified O4 bioconjugate against *E. coli* O4 isolates with characterized *gtrS* status as described in Example 4; the following isolates were *gtrS*-negative: A2625, stGVXN4988, OC24784, OC24787 and OC24788; the following isolates were *gtrS*-positive: Y1382, E551, OC24334, stGVXN4983, stGVXN4994 and OC24794; the negative control strain OC9487 (ATCC 35383; serotype O75) was also included;

[0049] **FIG. 3** shows Western blots of LPS extracted from *gtrS*-positive and –negative O4 isolates probed with pooled sera from rabbits immunized with modified O4 polysaccharide;

[0050] **FIGS. 4A and 4B** show antibody responses induced by glucosylated O4 (O4-Glc+)-EPA bioconjugates; **FIG. 4A** shows serum antibody levels measured by ELISA at day 0, 14 and 42 post-immunization; individual titers (log₁₀ EC50 titer) and GMT ± 95% CI are shown; the grey dotted line indicates the threshold above which the dilution curves of the samples have a 4PL fitting; **FIG. 4B** shows the results of the opsonophagocytic (OPK) assay to determine the functionality of the antibodies in serum samples obtained at day 42 post-immunization with glucosylated O4 (O4-Glc+)-EPA bioconjugate (4.0 µg); Wilcoxon rank sum test and Bonferroni correction; *P≤0.05, ***P≤0.0001;

[0051] **FIG. 5** shows the boost effect of glucosylated O4 (O4 Glc+)-EPA bioconjugate in Sprague Dawley rats immunized at 3 different doses as described in Example 4; serum antibody levels were measured by ELISA at day 0, 14 and 42 post-immunization; individual titers (log₁₀

EC50 titer) are shown for each animal; the lines between the data points connect IgG titers for each animal in time; the grey dotted line indicates the threshold above which the dilution curves of the samples have a 4PL fitting; statistical analysis was performed with Wilcoxon signed-rank test and Bonferroni correction for multiple comparisons (day 14 vs day 0, P = 0.012 for 4.0 µg/dose; day 42 vs day 0, P = 0.006 for all doses; day 42 vs day 14, P = 0.006 for all doses);

[0052] **FIG. 6** shows the functionality of antibodies induced by O4-Glc+-EPA bioconjugate; Sprague Dawley rats were immunized intramuscularly 3 times with formulation buffer or O4(Glc+)-EPA bioconjugate at 4.00 µg/dose; functionality of the antibodies was determined by opsonophagocytic killing assay (OPKA) using O4(Glc+) and O4(Glc-) *E. coli* strains; individual opsonic titers (OI) and GMT ± 95% CI are shown;

[0053] **FIG. 7** shows capillary electrophoresis readout of PglB screen visualizing O4-Glc+ bioconjugate production for each tested strain in a blot-like image, using monoclonal antibodies to detect O4-Glc+ bioconjugate in the periplasmic fraction. Mono-glycosylated product approximately 180 kDa, di-glycosylated product approximately 320 kDa and tri-glycosylated product approximately 450 kDa. A) First screening round. Wt PglB in lane 3, N311V-PglB in lanes 2 and 4, empty control strain in lane 1 and other PglB variants in lanes 5 and 6. B) Second screening round. N311V PglB in lane 3, N311V+Y77H PglB in lane 9, empty control strain in lanes 1 and 2, other PglB variants in remaining lanes.

[0054] **FIG. 8** shows antibody responses induced by ExPEC10V vaccine in New Zealand White rabbits. Animals received 3 intramuscular immunizations with ExPEC10V or saline administered 2 weeks apart. ExPEC10V vaccine was administered at 3 different concentrations (group 1: high dose, group 2: medium dose and group 3: low dose, Table 11) and a control group received only saline (group 4, 0.9% (w/v) sodium chloride solution). Antibody levels were measured by ELISA at day 0 (pre-vaccination) and days 14, 27 and 42 (post-vaccination). Individual titers (EC50 titer) and geometric mean titers (GMT) ± 95% CI are shown. Wilcoxon Rank Sum test with Bonferroni correction for multiple comparisons. Comparisons ExPEC10V vaccinated animals (group 1, 2 and 3) versus saline control (group 4). *p ≤ 0.05, **p ≤ 0.01; ***p ≤ 0.001; ****p ≤ 0.0001. LOD: limit of detection.

[0055] **FIG. 9** shows antibody responses induced by ExPEC10V. New Zealand White rabbits received 3 intramuscular immunizations with ExPEC10V (105.6 µg total polysaccharide) or

0.9% w/v sodium chloride solution (control). IgG titers were determined by ELISA at day 1 (pre-immunization, n = 20/group), day 31 (post-immunization, n = 20/group) and day 50 (post-immunization, n = 10/group). Plots show individual titers and geometric mean ± 95% confidence interval for each group. Differences in IgG titers between the ExPEC10V and control group were analyzed using a Tobit model with a likelihood ratio test. P-values ≤ 0.05 were considered significant. *P ≤ 0.05, ****P ≤ 0.0001.

[0056] **FIG. 10** shows the overall study design for a phase 1/2a clinical trial with ExPEC10V vaccine in humans. **FIG. 10A** shows the overall study design for Cohort 1, and **FIG. 10B** shows the overall study design for Cohort 2. See Example 11 for details.

DETAILED DESCRIPTION OF THE INVENTION

[0057] Various publications, articles and patents are cited or described in the background and throughout the specification; each of these references is herein incorporated by reference in its entirety. Discussion of documents, acts, materials, devices, articles or the like which has been included in the present specification is for the purpose of providing context for the invention. Such discussion is not an admission that any or all of these matters form part of the prior art with respect to any inventions disclosed or claimed.

[0058] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention pertains. Otherwise, certain terms used herein have the meanings as set forth in the specification.

[0059] It must be noted that as used herein and in the appended claims, the singular forms “a,” “an,” and “the” include plural reference unless the context clearly dictates otherwise.

[0060] Unless otherwise indicated, the term “at least” preceding a series of elements is to be understood to refer to every element in the series.

[0061] Those skilled in the art will recognize or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the invention.

[0062] Throughout this specification and the claims which follow, unless the context requires otherwise, the word “comprise”, and variations such as “comprises” and “comprising”, will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not

the exclusion of any other integer or step or group of integer or step. When used herein the term “comprising” can be substituted with the term “containing” or “including” or sometimes when used herein with the term “having”.

[0063] When used herein “consisting of” excludes any element, step, or ingredient not specified in the claim element. When used herein, “consisting essentially of” does not exclude materials or steps that do not materially affect the basic and novel characteristics of the claim. Any of the aforementioned terms of “comprising,” “containing,” “including,” and “having,” whenever used herein in the context of an aspect or embodiment of the invention can be replaced with the term “consisting of” or “consisting essentially of” to vary scopes of the disclosure.

[0064] As used herein, the conjunctive term “and/or” between multiple recited elements is understood as encompassing both individual and combined options. For instance, where two elements are conjoined by “and/or,” a first option refers to the applicability of the first element without the second. A second option refers to the applicability of the second element without the first. A third option refers to the applicability of the first and second elements together. Any one of these options is understood to fall within the meaning, and therefore satisfy the requirement of the term “and/or” as used herein. Concurrent applicability of more than one of the options is also understood to fall within the meaning, and therefore satisfy the requirement of the term “and/or.”

[0065] The identification of an O-antigen structural modification, namely glucose branching, within the *E. coli* O4 serotype (Jann et al., 1993) presents a challenge to the discovery and development of a glycoconjugate vaccine targeting bacterial isolates within this serotype. The proportion of clinical contemporary O4 isolates expressing the unmodified (not having a glucose side-branch) and modified (having a glucose side-branch) forms of the O4 O-antigen is unknown. Obtaining information on this characteristic is critical for selecting the relevant antigenic structure. In addition, the extent to which vaccine induced antibodies elicited to one form of the O4 polysaccharide will cross-react with the other form has not been determined. Purification of O-antigen free from lipid A and subsequent chemical conjugation to a carrier protein is a lengthy and laborious process. Additionally, the purification, lipid A detoxification and chemical conjugation processes can result in loss of epitopes, antigen heterogeneity and reduced immunogenicity of the conjugated polysaccharide. Synthesis of glycoconjugates by bioconjugation can overcome these limitations of classical purification and chemical

conjugation, but the *in vivo* synthesis of glucose-branched O4 O-antigen requires the activity of a polysaccharide branching enzyme, which lies outside of the *rfb* gene cluster. To date, the O-antigen modifying enzyme responsible for glucose-branching in O4 *E. coli* strains has not been identified. Cloning the O4 *rfb* gene cluster into the bioconjugation *E. coli* strain expressing PglB will not be sufficient to synthesize the glucose-branched O4 glycoconjugate, but rather would only produce non-glucose-branched O4 bioconjugates (the structure of the glycan thereof is shown in Formula (O4) in Table 1). As used herein, the terms “glucosylated O4”, “glucose-branched O4”, “O4 Glc+” and “Glc+ O4” O-antigen refer to O4 O-antigen with a glucose side-branch, and the structure thereof is shown in formula (O4-Glc+) in Table 1.

[0066] Disclosed herein is the gene encoding the O-antigen modifying enzyme responsible for glucose branching of the *E. coli* O4 antigen polysaccharide. Also disclosed herein are host cells, e.g., recombinantly engineered host cells comprising nucleic acid encoding enzymes capable of producing bioconjugates comprising the glucosylated O4 antigen polysaccharide covalently bound to a carrier protein *in vivo*. Such host cells can be used to generate bioconjugates comprising the glucosylated O4 antigen linked to a carrier protein, which can be used in, e.g., the formulation of therapeutic and/or prophylactic compositions (e.g., vaccines). Further provided herein are compositions comprising bioconjugates of the glucosylated O4 antigen polysaccharide, alone or in combination with other *E. coli* antigens (e.g., O1, O2, O6, O8, O15, O16, O18, O25, and/or O75 antigen polysaccharides and subserotypes thereof). The compositions can be used in prophylactic and/or therapeutic methods, e.g., vaccination of hosts against infection with *E. coli*, and are useful in the generation of antibodies, which can be used, e.g., in therapeutic methods such as for immunization of subjects.

[0067] As used here, the terms “O-antigen,” “O-antigen polysaccharide,” “O-antigen saccharide,” and “OPS” refer to the O-antigen of Gram-negative bacteria. Typically, an O-antigen is a polymer of immunogenic repeating polysaccharide units. In a particular embodiment, the terms “O-antigen,” “O-antigen polysaccharide,” and “OPS” refer to the O-antigen of *Escherichia coli*. Different serotypes of *E. coli* express different O-antigens. In *E. coli*, the gene products involved in O-antigen biogenesis are encoded by the *rfb* gene cluster.

[0068] As used herein, “*rfb* cluster” and “*rfb* gene cluster” refer to a gene cluster that encodes enzymatic machinery capable of synthesizing an O-antigen backbone structure. The term *rfb*

cluster can apply to any O-antigen biosynthetic cluster, and preferably refers to a gene cluster from the genus *Escherichia*, particularly *E. coli*.

[0069] As used herein, the term “O1A” refers to the O1A antigen of *E. coli* (a subserotype of *E. coli* serotype O1). The term “O2” refers to the O2 antigen of *E. coli* (*E. coli* serotype O2). The term “O6A” refers to the O6A antigen of *E. coli* (a subserotype of *E. coli* serotype O6). The term “O8” refers to the O8 antigen of *E. coli* (*E. coli* serotype O8). The term “O15” refers to the O15 antigen of *E. coli* (*E. coli* serotype O15). The term “O16” refers to the O16 antigen of *E. coli* (*E. coli* serotype O16). The term “O18A” refers to the O18A antigen of *E. coli* (a subserotype of *E. coli* serotype O18). The term “O25B” refers to the O25B antigen from *E. coli* (a subserotype of *E. coli* serotype O25). The term “O75” refers to the O75 antigen of *E. coli* (*E. coli* serotype O75).

[0070] The structures of *E. coli* O-antigen polysaccharides referred to throughout this application are shown below in Table 1. A single repeating unit for each *E. coli* O-antigen polysaccharide is shown.

Table 1: Structures of *E. coli* O-antigen Polysaccharides

<i>E. coli</i> O-antigen Polysaccharide	Structure of Repeating Unit¹
Non-glucosylated O4 antigen polysaccharide (O4-Glc-)	$[\rightarrow 2) - \alpha-L-Rhap-(1 \rightarrow 6) - \alpha-D-Glcp-(1 \rightarrow 3) - \alpha-L-FucpNAc-(1 \rightarrow 3) - \beta-D-GlcpNAc-(1 \rightarrow]_n$
Glucosylated O4 antigen polysaccharide (O4-Glc+)	$\alpha-D-Glcp$ 1 ↓ 3 $[\rightarrow 2) - \alpha-L-Rhap-(1 \rightarrow 6) - \alpha-D-Glcp-(1 \rightarrow 3) - \alpha-L-FucpNAc-(1 \rightarrow 3) - \beta-D-GlcpNAc-(1 \rightarrow]_n$
O1A antigen polysaccharide (O1A)	$[\rightarrow 3) - \alpha-L-Rhap-(1 \rightarrow 3) - \alpha-L-Rhap-(1 \rightarrow 3) - \beta-L-Rhap-(1 \rightarrow 4) - \beta-D-GlcpNAc-(1 \rightarrow]_n$ 2 ↑ 1 $\beta-D-ManpNAc$

O2 antigen polysaccharide (O2)	$[\rightarrow 3)\alpha-L-Rhap-(1\rightarrow 2)\alpha-L-Rhap-(1\rightarrow 3)\beta-L-Rhap-(1\rightarrow 4)\beta-D-GlcNAc-(1\rightarrow]_n$ 2 ↑ 1 $\alpha-D-Fucp3NAc$
O6A antigen polysaccharide (O6)	$[\rightarrow 4)\alpha-D-GalpNAc-(1\rightarrow 3)\beta-D-Manp-(1\rightarrow 4)\beta-D-Manp-(1\rightarrow 3)\alpha-D-GlcNAc-(1\rightarrow]_n$ 2 ↑ 1 $\beta-D-GlcP$
O8 antigen polysaccharide (O8)	$\alpha-D-Manp3Me-(1\rightarrow [3)\beta-D-Manp-(1\rightarrow 2)\alpha-D-Manp-(1\rightarrow 2)\alpha-D-Manp-(1\rightarrow]_n$
O15 antigen polysaccharide (O15)	$[\rightarrow 2)\beta-D-Galp-(1\rightarrow 3)\alpha-L-FucpNAc-(1\rightarrow 3)\beta-D-GlcNAc-(1\rightarrow]_n$
O16 antigen polysaccharide (O16)	$[\rightarrow 2)\beta-D-Galf-(1\rightarrow 6)\alpha-D-GlcP-(1\rightarrow 3)\alpha-L-Rhap-(1\rightarrow 3)\alpha-D-GlcNAc-(1\rightarrow]_n$ 2 ↑ Ac

O18A antigen polysaccharide (O18A)	$[\rightarrow 2)\text{-}\alpha\text{-L-Rhap-(1}\rightarrow 6)\text{-}\alpha\text{-D-Glc}\beta\text{-(1}\rightarrow 4)\text{-}\alpha\text{-D-Gal}\beta\text{-(1}\rightarrow 3)\text{-}\alpha\text{-D-Glc}\beta\text{pNAc-(1}\rightarrow]_n$
O25B antigen polysaccharide (O25B)	$\begin{array}{c} \beta\text{-D-Glc}\beta \\ \\ 1 \\ \downarrow \\ 6 \\ [\rightarrow 4)\text{-}\alpha\text{-D-Glc}\beta\text{-(1}\rightarrow 3)\text{-}\alpha\text{-L-Rhap-(1}\rightarrow 3)\text{-}\beta\text{-D-Glc}\beta\text{pNAc-(1}\rightarrow]_n \\ \qquad \qquad \\ 3 \qquad \qquad 2 \\ \uparrow \qquad \qquad \uparrow \\ 1 \qquad \qquad \text{Ac} \\ \alpha\text{-L-Rhap} \end{array}$
O75 antigen polysaccharide (O75)	$\begin{array}{c} \beta\text{-D-Man}\beta \\ \\ 1 \\ \downarrow \\ 4 \\ [\rightarrow 3)\text{-}\alpha\text{-D-Gal}\beta\text{-(1}\rightarrow 4)\text{-}\alpha\text{-L-Rhap-(1}\rightarrow 3)\text{-}\beta\text{-D-Glc}\beta\text{pNAc-(1}\rightarrow]_n \end{array}$

¹ Each n is independently an integer of 1 to 100, such as 1-50, 1-40, 1-30, 1-20, and 1-10, 3-50, 3-40, e.g. at least 5, such as 5-40, e.g. 7-30, e.g. 7 to 25, e.g. 10 to 20, but in some instances can be 1-2.

[0071] All monosaccharides described herein have their common meaning known in the art. Monosaccharides can have the D or L configuration. If D or L is not specified, the sugar is understood to have the D configuration. Monosaccharides are typically referred to by abbreviations commonly known and used in the art. For example, Glc refers to glucose; D-Glc refers to D-glucose; and L-Glc refers to L-glucose. Other common abbreviations for monosaccharides include: Rha, rhamnose; GlcNAc, N-acetylglucosamine; GalNAc, N-acetylgalactosamine; Fuc, fucose; Man, mannose; Man3Me, 3-O-methyl-mannose; Gal, galactose; FucNAc, N-acetylglucosamine; and Rib, ribose. The suffix “f” refers to furanose and the suffix “p” refers to pyranose.

[0072] The terms “RU,” “repeat unit,” and ”repeating unit” as used with respect to an O-antigen refer to the biological repeat unit (BRU) of an O-antigen as it is synthesized *in vivo* by cellular machinery (e.g., glycosyltransferases). The number of RUs of an O-antigen may vary per serotype, and in embodiments of the invention typically varies from about 1-100 RUs, preferably about 1 to 50 RUs, such as 1-50 RUs, 1-40 RUs, 1-30 RUs, 1-20 RUs, and 1-10 RUs, and more preferably at least 3 RUs, at least 4 RUs, at least 5 RUs, such as 3-50 RUs, preferably 5-40 RUs, e.g. 7-25 RUs, e.g. 10-20 RUs. However, in some instances, the number of RUs of an O-antigen can be 1-2. The structure of each O-antigen that is specifically described herein is shown containing one RU with the variable “n” designating the number of RUs. In each O-antigen polysaccharide in a bioconjugate of the invention, n is independently an integer of 1-100, such as 1-50, 1-40, 1-30, 1-20, 1-10, preferably at least 3, more preferably at least 5, such as 3-50, preferably 5-40 (e.g. 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40), but in some instances can be 1-2. In some embodiments n is independently an integer of about 7-25, e.g. about 10-20. The values may vary between individual O-antigen polysaccharides in a composition, and are provided here as average values, i.e. if a bioconjugate is described herein as having an n that is independently an integer of 5-40, the composition contains a majority of O-antigen polysaccharides with 5-40 repeat units, but may also contain some O-antigen polysaccharides that have less than 5 repeat units or more than 40 repeat units.

[0073] The term “glycoconjugate” refers to a sugar or saccharide antigen (e.g., oligo- and polysaccharide)-protein conjugate linked to another chemical species, including but not limited to proteins, peptides, lipids, etc. Glycoconjugates can be prepared chemically, e.g., by chemical (synthetic) linkage of the protein and sugar or saccharide antigen. The term glycoconjugate also includes bioconjugates.

[0074] The term “bioconjugate” refers to a conjugate between a protein (e.g., a carrier protein) and a sugar or saccharide antigen (e.g., oligo- and polysaccharide) prepared in a host cell background, preferably a bacterial host cell, e.g. an *E.coli* host cell, wherein host cell machinery links the antigen to the protein (e.g., N-links). Preferably, the term “bioconjugate” refers to a conjugate between a protein (e.g., carrier protein) and an O-antigen, preferably an *E. coli* O-antigen (e.g., O1A, O2, glucosylated O4, O6A, O8, O15, O16, O18A, O25B, O75, etc.) prepared in a host cell background, wherein host cell machinery links the antigen to the protein (e.g., N-links). Because bioconjugates are prepared in host cells by host cell machinery, the antigen and protein are covalently linked via a glycosidic linkage or bond in a bioconjugate. Bioconjugates can be prepared in recombinant host cells engineered to express the cellular machinery needed to synthesize the O-antigen and/or link the O-antigen to the target protein. Bioconjugates, as described herein, have advantageous properties over chemically prepared glycoconjugates where the glycans are purified from bacterial cell walls and subsequently chemically coupled to a carrier protein, e.g., bioconjugates require fewer chemicals in manufacture and are more consistent in terms of the final product generated, and contain less or no free (i.e. unbound to carrier protein) glycan. Thus, in typical embodiments, bioconjugates are preferred over chemically produced glycoconjugates.

[0075] The term “about,” when used in conjunction with a number, refers to any number within ± 1 , ± 5 or $\pm 10\%$ of the referenced number.

[0076] The term “percent (%) sequence identity” or “% identity” describes the number of matches (“hits”) of identical amino acids of two or more aligned amino acid sequences as compared to the number of amino acid residues making up the overall length of the amino acid sequences. In other terms, using an alignment, for two or more sequences the percentage of amino acid residues that are the same (e.g. 90%, 95%, 97% or 98% identity) may be determined, when the sequences are compared and aligned for maximum correspondence as measured using

a sequence comparison algorithm as known in the art, or when manually aligned and visually inspected. The sequences which are compared to determine sequence identity may thus differ by substitution(s), addition(s) or deletion(s) of amino acids. Suitable programs for aligning protein sequences are known to the skilled person. The percentage sequence identity of protein sequences can, for example, be determined with programs such as CLUSTALW, Clustal Omega, FASTA or BLAST, e.g using the NCBI BLAST algorithm (Altschul SF, et al (1997), Nucleic Acids Res. 25:3389-3402).

[0077] For example, for amino acid sequences, sequence identity and/or similarity can be determined by using standard techniques known in the art, including, but not limited to, the local sequence identity algorithm of Smith and Waterman, 1981, Adv. Appl. Math. 2:482, the sequence identity alignment algorithm of Needleman and Wunsch, 1970, J. Mol. Biol. 48:443, the search for similarity method of Pearson and Lipman, 1988, Proc. Natl. Acad. Sci. U.S.A. 85:2444, computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Drive, Madison, Wis.), the Best Fit sequence program described by Devereux et al, 1984, Nucl. Acid Res. 12:387-395, preferably using the default settings, or by inspection. In certain embodiments, percent identity is calculated by FastDB based upon the following parameters: mismatch penalty of 1; gap penalty of 1; gap size penalty of 0.33; and joining penalty of 30, "Current Methods in Sequence Comparison and Analysis," Macromolecule Sequencing and Synthesis, Selected Methods and Applications, pp 127-149 (1988), Alan R. Liss, Inc.

[0078] Another example of a useful algorithm is the BLAST algorithm, described in: Altschul et al, 1990, J. Mol. Biol. 215:403-410; Altschul et al, 1997, Nucleic Acids Res. 25:3389-3402; and Karin et al, 1993, Proc. Natl. Acad. Sci. U.S.A. 90:5873-5787. A particularly useful BLAST program is the WU-BLAST-2 program which was obtained from Altschul et al, 1996, Methods in Enzymology 266:460-480. WU-BLAST-2 uses several search parameters, most of which are set to the default values.

[0079] An additional useful algorithm is gapped BLAST as reported by Altschul et al, 1993, Nucl. Acids Res. 25:3389-3402.

[0080] The term "Invasive Extraintestinal pathogenic *Escherichia coli* (ExPEC) disease (IED)" is defined herein as an acute illness consistent with systemic bacterial infection, which is

microbiologically confirmed either by the isolation and identification of *E. coli* from blood or other normally sterile body sites, or by the isolation and identification of *E. coli* from urine in a patient with presence of signs and symptoms of invasive disease (systemic inflammatory response syndrome (SIRS), sepsis or septic shock) and no other identifiable source of infection.

[0081] Bioconjugates of *E. coli* glucosylated O4 Antigen Polysaccharides

[0082] In one aspect, provided herein is a bioconjugate of an *E. coli* glucosylated O4 antigen polysaccharide covalently linked to a carrier protein. As used herein, the term “O4” refers to the O4 antigen from *E. coli* (*E. coli* serotype O4). O-antigen structural modification is known to exist within the *E. coli* O4 serotype. In particular, some O4 serotypes express a modified O-antigen having a branched glucose unit. As used herein, “glucosylated O4 antigen,” “glucosylated O4 antigen polysaccharide,” “O4-Glc+ antigen polysaccharide,” and “O4-Glc+ antigen” refer to an O4 antigen (e.g., *E. coli* O4 antigen) having a glucose branch, in which D-glucose is linked to L-rhamnose in the repeating unit L-Rha→D-Glc→L-FucNAc→D-GlcNAc. In a particular embodiment, an *E. coli* glucosylated O4 antigen polysaccharide comprises the structure of formula (O4-Glc+), as shown in Table 1, wherein n is an integer of 1 to 100. In preferred embodiments, n is an integer of 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20.

[0083] *E. coli* O4 strains, independent of glucose branching status, carry a substantially identical *rfb* gene cluster encoding the genes responsible for production of the O4 antigen polysaccharide. However, *in vivo* synthesis of the modified O4 antigen having glucose branching requires the activity of a polysaccharide branching enzyme, which lies outside of the *rfb* gene cluster. To the best of the knowledge of the inventors, the identity of the polysaccharide branching enzyme responsible for glucose modification of the O4 antigen has remained unknown to date. Here, the inventors discovered the sequence of the polysaccharide branching enzyme responsible for glucose modification of the O4 antigen. Identification of this enzyme enables production of bioconjugates of the modified O4 antigen polysaccharide having glucose branching. The glucose modified form of the O4 antigen polysaccharide is present in predominant serotypes and can thus be used to provide an improved immune response, e.g for prophylactic or therapeutic use.

[0084] In particular, provided herein is the sequence of a *gtrS* gene encoding a glucosyltransferase enzyme specific for *E. coli* serotype O4 that glucosylates the O4 antigen. In

general, the *gtrA*, *gtrB*, and *gtrS* genes encodes the enzymes responsible for O-antigen glucosylation. While the *gtrA* and *gtrB* genes in different serotypes are highly homologous and interchangeable, the *gtrS* gene encodes a serotype specific O-antigen glucosyl transferase. The *gtrS* gene of *E. coli* serotype O4 encodes the GtrS enzyme that modifies the O4 antigen by introducing glucose branching. Characterization of contemporary clinical *E. coli* isolates of the O4 serotype revealed the presence of *gtrS* in 78% of tested isolates, indicating that *E. coli* O4 antigen polysaccharide modified with the addition of a glucose residue is predominant in current infecting isolates.

[0085] In one embodiment, provided herein is a nucleic acid of a *gtrS* gene from *E. coli* serotype O4 encoding a GtrS glucosyltransferase comprising the amino acid sequence of SEQ ID NO: 4. In another embodiment, a *gtrS* nucleic acid encodes a GtrS protein from *E. coli* serotype O4 that is about 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of SEQ ID NO: 4, preferably 98%, 99%, or 100% identical to the amino acid sequence of SEQ ID NO: 4. A GtrS protein that is at least 80% identical to the amino acid sequence of SEQ ID NO: 4 is capable of specifically glucosylating the *E. coli* O4 antigen polysaccharide to obtain a glucosylated O4 antigen having the structure of Formula (O4-Glc+) as shown in Table 1. One of ordinary skill in the art will be able to make mutated forms of the GtrS protein of SEQ ID NO: 4 having at least 80% sequence identity to SEQ ID NO: 4, and test such sequences for glucosylation activity of the *E. coli* O4 antigen in view of the present disclosure. Recombinant host cells comprising nucleic acid sequence encoding the glucosyl transferase *gtrS* gene of *E. coli* serotype O4, and use of the recombinant host cells in production of the glucose modified O4 antigen polysaccharides and bioconjugates thereof are described in greater detail below.

[0086] Sequences for *gtrA* and *gtrB* encoded proteins, which function as bactoprenol-linked glucose translocase (GtrA, flips the bactoprenol-linked glucose over the inner membrane to the periplasm) and bactoprenol glucosyl transferase (GtrB, links glucose to bactoprenol), respectively, may comprise amino acid sequences that are at least about 80% identical to SEQ ID NOs: 7 and 8, respectively. In certain embodiments, nucleic acid sequences encoding GtrA and GtrB proteins that are at least about 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NOs: 7 and 8, respectively, and having bactoprenol-linked glucose

translocase and bactoprenol glucosyl transferase activity, respectively, are also present in the host cells of the invention, that further comprise an O4-specific *rfb* locus, the O4-specific GtrS encoding sequence described above, an oligosaccharyl transferase as described herein, and a sequence encoding a carrier protein having one or more glycosylation consensus sequences as described herein, to produce bioconjugates of *E.coli* glucosylated O4 serotype (comprising glycan structure of Formula (O4-Glc+) in Table 1).

[0087] Bioconjugates of an *E. coli* glucosylated O4 antigen polysaccharide provided herein are covalently linked to a carrier protein, preferably by a glycosidic linkage. Any carrier protein known to those skilled in the art in view of the present disclosure can be used. Suitable carrier proteins include, but are not limited to, detoxified Exotoxin A of *P. aeruginosa* (EPA), *E. coli* flagellin (FliC), CRM197, maltose binding protein (MBP), Diphtheria toxoid, Tetanus toxoid, detoxified hemolysin A of *S. aureus*, clumping factor A, clumping factor B, *E. coli* heat labile enterotoxin, detoxified variants of *E. coli* heat labile enterotoxin, Cholera toxin B subunit (CTB), cholera toxin, detoxified variants of cholera toxin, *E. coli* Sat protein, the passenger domain of *E. coli* Sat protein, *Streptococcus pneumoniae* Pneumolysin, Keyhole limpet hemocyanin (KLH), *P. aeruginosa* PcrV, outer membrane protein of *Neisseria meningitidis* (OMPC), and protein D from non-typeable *Haemophilus influenzae*. Bioconjugation with various different carrier proteins containing the required consensus glycosylation sequence has been described, showing that a wide range of proteins can be glycosylated using this technology (see, e.g. WO 06/119987, WO 2015/124769, WO 2015/158403, WO 2015/82571, WO 2017/216286, and WO 2017/67964, together showing a wide variety of carrier proteins that were successfully used in bioconjugation).

[0088] In certain embodiments a carrier protein is modified, e.g., modified in such a way that the protein is less toxic and/or more susceptible to glycosylation. In a specific embodiment, the carrier proteins used herein are modified such that the number of glycosylation sites in the carrier proteins is maximized in a manner that allows for lower concentrations of the protein to be administered, e.g., in an immunogenic composition, particularly in its bioconjugate form.

[0089] Thus, in certain embodiments, the carrier proteins described herein are modified to include 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more glycosylation sites than would normally be associated with the carrier protein (e.g., relative to the number of glycosylation sites associated with the

carrier protein in its native/natural, i.e., “wild-type” state). Introduction of glycosylation sites into a carrier protein can be accomplished by insertion of a glycosylation consensus sequence anywhere in the primary structure of the protein by, e.g., adding new amino acids to the primary structure of the protein such that a glycosylation site is added in full or in part, or by mutating existing amino acids in the protein in order to generate a glycosylation site. One of ordinary skill in the art will recognize that the amino acid sequence of a protein can be readily modified using approaches known in the art, e.g., recombinant approaches that include modification of the nucleic acid sequence encoding the protein. In specific embodiments, glycosylation consensus sequences are introduced into specific regions of the carrier protein, e.g., surface structures of the protein, at the N or C termini of the protein, and/or in loops that are stabilized by disulfide bridges at the base of the protein. In some embodiments, a glycosylation consensus sequence can be extended by addition of lysine residues for more efficient glycosylation.

[0090] Exemplary examples of glycosylation consensus sequences that can be inserted into or generated in a carrier protein include Asn-X-Ser(Thr), wherein X can be any amino acid except Pro (SEQ ID NO: 1); and Asp(Glu)-X-Asn-Z-Ser(Thr), wherein X and Z are independently selected from any amino acid except Pro (SEQ ID NO: 2).

[0091] In some embodiments, the *E. coli* glucosylated O4 antigen polysaccharide is covalently linked to an asparagine (Asn) residue in the carrier protein (e.g., N-linked), wherein the Asn residue is present in a glycosylation site comprising a glycosylation consensus sequence having SEQ ID NO: 1, more preferably having SEQ ID NO: 2. Typically, a carrier protein comprises 1-10 glycosylation sites, preferably 2 to 4 glycosylation sites, most preferably 4 glycosylation sites, such as 1-10, preferably 2-4, and more preferably 4 glycosylation sites each comprising a glycosylation consensus sequences having the amino acid sequence of SEQ ID NO: 1, and more preferably the amino acid sequence of SEQ ID NO: 2.

[0092] In particular embodiments, a carrier protein is a detoxified Exotoxin A of *P. aeruginosa*. For EPA, various detoxified protein variants have been described in literature and could be used as carrier proteins. For example, detoxification can be achieved by mutating and deleting the catalytically essential residues L552V and ΔE553 according to Lukac et al., 1988, *Infect Immun*, 56: 3095-3098, and Ho et al., 2006, *Hum Vaccin*, 2:89–98. As used herein, “EPA” refers to a detoxified Exotoxin A of *P. aeruginosa*. In those embodiments, wherein the carrier

protein is EPA, an *E. coli* glucosylated O4 antigen polysaccharide can be covalently linked to an Asn residue in a glycosylation site comprising a glycosylation consensus sequence having SEQ ID NO: 1, and preferably covalently linked to an Asn residue in a glycosylation site comprising a glycosylation consensus sequence having SEQ ID NO: 2. Preferably, the EPA carrier protein comprises 1-10 glycosylation sites, preferably 2 to 4 glycosylation sites, most preferably 4 glycosylation sites, such as 1-10, preferably 2-4, and more preferably 4 glycosylation sites each comprising a glycosylation consensus sequence having the amino acid sequence of SEQ ID NO: 1, and more preferably the amino acid sequence of SEQ ID NO: 2.

[0093] In some embodiments, the EPA carrier protein comprises four glycosylation sites each comprising a glycosylation consensus sequence, for instance a glycosylation site comprising a glycosylation consensus sequence having SEQ ID NO: 2. As used herein, “EPA-4 carrier protein” and “EPA-4” refer to a detoxified Exotoxin A of *P. aeruginosa* carrier protein comprising four glycosylation sites each comprising a glycosylation consensus sequences having SEQ ID NO: 2. An exemplary preferred example of an EPA-4 carrier protein is EPA carrier protein comprising the amino acid sequence of SEQ ID NO: 3.

[0094] Compositions

[0095] In another aspect, provided herein is a composition comprising a bioconjugate of an *E. coli* glucosylated O4 antigen polysaccharide covalently linked to a carrier protein. The compositions provided herein can include any bioconjugate of an *E. coli* glucosylated O4 antigen polysaccharide covalently linked to a carrier protein (e.g., EPA) described herein.

[0096] In some embodiments, a composition is an immunogenic composition. As used herein, an “immunogenic composition” refers to a composition that can elicit an immune response in a host or subject to whom the composition is administered. Immunogenic compositions can further comprise a pharmaceutically acceptable carrier. In some embodiments, a composition is a pharmaceutical composition further comprising a pharmaceutically acceptable carrier. As used herein, a “pharmaceutically acceptable carrier” refers to a diluent, adjuvant, excipient or vehicle with which a composition is administered, and that is non-toxic and should not interfere with the efficacy of the active ingredient. For example, saline solutions and aqueous dextrose and glycerol solutions can also be employed as liquid carriers, particularly for injectable solutions. Suitable excipients include starch, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica

gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propylene, glycol, water, ethanol and the like. Other examples of suitable pharmaceutically acceptable carriers are described in “Remington's Pharmaceutical Sciences” by E.W. Martin.

[0097] In one embodiment, a composition of the invention comprises the bioconjugates of the invention in a Tris-buffered saline (TBS) pH 7.4 (e.g. containing Tris, NaCl and KCl, e.g. at 25 mM, 137 mM and 2.7 mM, respectively). In other embodiments, the compositions of the invention comprise bioconjugates of the invention in about 10 mM KH₂PO₄/Na₂HPO₄ buffer at pH of about 7.0, about 5% (w/v) sorbitol, about 10 mM methionine, and about 0.02% (w/v) polysorbate 80. In other embodiments, the compositions of the invention comprise bioconjugates of the invention in about 10 mM KH₂PO₄/Na₂HPO₄ buffer at pH of about 7.0, about 8% (w/v) sucrose, about 1 mM EDTA, and about 0.02% (w/v) polysorbate 80 (see e.g. WO 2018/077853 for suitable buffers for bioconjugates of *E.coli* O-antigens covalently bound to EPA carrier protein).

[0098] In some embodiments, the compositions described herein are monovalent formulations, and contain one *E. coli* O-antigen polysaccharide, e.g., in isolated form or as part of a glycoconjugate or bioconjugate, such as the *E. coli* glucosylated O4 antigen polysaccharide. Also provided herein are compositions (e.g., pharmaceutical and/or immunogenic compositions) that are multivalent compositions, e.g., bivalent, trivalent, tetravalent, etc. compositions. For example, a multivalent composition comprises more than one antigen, such as an *E. coli* O-antigen, glycoconjugate, or bioconjugate thereof. In particular embodiments, multivalent compositions provided herein comprise a bioconjugate of an *E. coli* glucosylated O4 antigen polysaccharide, and at least one additional antigen.

[0099] In one embodiment, a composition (e.g., pharmaceutical and/or immunogenic composition) is a monovalent composition comprising a biconjugate of an *E. coli* glucosylated O4 antigen polysaccharide covalently linked to a carrier protein as described herein.

[0100] In another embodiment, a composition (e.g., pharmaceutical and/or immunogenic composition) is a multivalent composition comprising an *E. coli* glucosylated O4 antigen polysaccharide covalently linked to a carrier protein as described herein, and at least one additional antigen.

[00101] In some embodiments, the additional antigen is antigen saccharide or polysaccharide, more preferably an *E. coli* O-antigen polysaccharide, such as *E. coli* O-antigens of one or more of the O1, O2, O6, O8, O15, O16, O18, O25, and O75 serotypes and subserotypes thereof. In some embodiments, each of the additional *E. coli* O-antigen polysaccharides is a glycoconjugate, meaning that the *E. coli* O-antigen polysaccharide is covalently linked to another chemical species, e.g., protein, peptide, lipid, etc., most preferably a carrier protein, such as by chemical or enzymatic methods. In preferred embodiments, each of the additional *E. coli* O-antigen polysaccharides is a bioconjugate in which the O-antigen polysaccharide is covalently linked to, e.g. a carrier protein, via a glycosidic bond enzymatically by host cell machinery. Compositions provided herein in certain embodiments can comprise 1-20 additional glycoconjugates, more preferably bioconjugates of *E. coli* O-antigen polysaccharides, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 additional glycoconjugates or preferably bioconjugates of *E. coli* O-antigen polysaccharides. Other antigens can be included in the compositions provided herein, such as peptide, protein, or lipid antigens, etc.

[00102] In some embodiments, a composition (e.g., pharmaceutical and/or immunogenic composition) comprises a bioconjugate of an *E. coli* glucosylated O4 antigen polysaccharide, and at least one additional antigen polysaccharide selected from the group consisting of *E. coli* O1A antigen polysaccharide, *E. coli* O2 antigen polysaccharide, *E. coli* O6A antigen polysaccharide, *E. coli* O8 antigen polysaccharide, *E. coli* O15 antigen polysaccharide, *E. coli* O16 antigen polysaccharide, *E. coli* O18A antigen polysaccharide, *E. coli* O25B antigen polysaccharide, and *E. coli* O75 antigen polysaccharide. Preferably, each of the additional O-antigen polysaccharides is covalently linked to a carrier protein, and is more preferably a bioconjugate.

[00103] In one embodiment, an O1A antigen polysaccharide (e.g., in isolated form or as part of a glycoconjugate or bioconjugate) is used in a composition provided herein (e.g., in combination with a glucosylated O4 antigen polysaccharide or bioconjugate thereof). In a specific embodiment, the O1A antigen polysaccharide comprises the structure of formula (O1A) as shown in Table 1, wherein n is an integer of 1-100, preferably 3-50, e.g. 5-40, e.g. 7 to 25, e.g. 10 to 20. Preferably, the O1A antigen polysaccharide is part of a bioconjugate and is covalently linked to a carrier protein, e.g., EPA.

[00104] In one embodiment, an O2 antigen polysaccharide (e.g., in isolated form or as part of a glycoconjugate or bioconjugate) is used in a composition provided herein (e.g., in combination with a glucosylated O4 antigen polysaccharide or bioconjugate thereof). In a specific embodiment, the O2 antigen polysaccharide comprises the structure of formula (O2) as shown in Table 1, wherein n is an integer of 1-100, preferably 3-50, e.g. 5-40, e.g. 7 to 25, e.g. 10 to 20. Preferably, the O2 antigen polysaccharide is part of a bioconjugate and is covalently linked to a carrier protein, e.g., EPA.

[00105] In one embodiment, an O6A antigen polysaccharide (e.g., in isolated form or as part of a glycoconjugate or bioconjugate) is used in a composition provided herein (e.g., in combination with a glucosylated O4 antigen polysaccharide or bioconjugate thereof). In a specific embodiment, the O6A antigen polysaccharide comprises the structure of formula (O6A) as shown in Table 1, wherein n is an integer of 1-100, preferably 3-50, e.g. 5-40, e.g. 7 to 25, e.g. 10 to 20. Preferably, the O6A antigen polysaccharide is part of a bioconjugate and is covalently linked to a carrier protein, e.g., EPA.

[00106] In one embodiment, an O8 antigen polysaccharide (e.g., in isolated form or as part of a glycoconjugate or bioconjugate) is used in a composition provided herein (e.g., in combination with a glucosylated O4 antigen polysaccharide or bioconjugate thereof). In a specific embodiment, the O8 antigen polysaccharide comprises the structure of formula (O8) as shown in Table 1, wherein n is an integer of 1-100, preferably 3-50, e.g. 5-40, e.g. 7 to 25, e.g. 10 to 20. Preferably, the O8 antigen polysaccharide is part of a bioconjugate and is covalently linked to a carrier protein, e.g., EPA.

[00107] In one embodiment, an O15 antigen polysaccharide (e.g., in isolated form or as part of a glycoconjugate or bioconjugate) is used in a composition provided herein (e.g., in combination with a glucosylated O4 antigen polysaccharide or bioconjugate thereof). In a specific embodiment, the O15 antigen polysaccharide comprises the structure of formula (O15) as shown in Table 1, wherein n is an integer of 1-100, preferably 3-50, e.g. 5-40, e.g. 7 to 25, e.g. 10 to 20. Preferably, the O15 antigen polysaccharide is part of a bioconjugate and is covalently linked to a carrier protein, e.g., EPA.

[00108] In one embodiment, an O16 antigen polysaccharide (e.g., in isolated form or as part of a glycoconjugate or bioconjugate) is used in a composition provided herein (e.g., in combination

with a glucosylated O4 antigen polysaccharide or bioconjugate thereof). In a specific embodiment, the O16 antigen polysaccharide comprises the structure of formula (O16) as shown in Table 1, wherein n is an integer of 1-100, preferably 3-50, e.g. 5-40, e.g. 7 to 25, e.g. 10 to 20. Preferably, the O16 antigen polysaccharide is part of a bioconjugate and is covalently linked to a carrier protein, e.g., EPA.

[00109] In one embodiment, an O18A antigen polysaccharide (e.g., in isolated form or as part of a glycoconjugate or bioconjugate) is used in a composition provided herein (e.g., in combination with a glucosylated O4 antigen polysaccharide or bioconjugate thereof). In a specific embodiment, the O18A antigen polysaccharide comprises the structure of formula (O18A) as shown in Table 1, wherein n is an integer of 1-100, preferably 3-50, e.g. 5-40, e.g. 7 to 25, e.g. 10 to 20. Preferably, the O18A antigen polysaccharide is part of a bioconjugate and is covalently linked to a carrier protein, e.g., EPA.

[00110] In one embodiment, an O25B antigen polysaccharide (e.g., in isolated form or as part of a glycoconjugate or bioconjugate) is used in a composition provided herein (e.g., in combination with a glucosylated O4 antigen polysaccharide or bioconjugate thereof). In a specific embodiment, the O25B antigen polysaccharide comprises the structure of formula (O25B) as shown in Table 1, wherein n is an integer of 1-100, preferably 3-50, e.g. 5-40, e.g. 7 to 25, e.g. 10 to 20. Preferably, the O25B antigen polysaccharide is part of a bioconjugate and is covalently linked to a carrier protein, e.g., EPA.

[00111] In one embodiment, an O75 antigen polysaccharide (e.g., in isolated form or as part of a glycoconjugate or bioconjugate) is used in a composition provided herein (e.g., in combination with a glucosylated O4 antigen polysaccharide or bioconjugate thereof). In a specific embodiment, the O75 antigen polysaccharide comprises the structure of formula (O75) as shown in Table 1, wherein n is an integer of 1-100, preferably 3-50, e.g. 5-40, e.g. 7 to 25, e.g. 10 to 20. Preferably, the O75 antigen polysaccharide is part of a bioconjugate and is covalently linked to a carrier protein, e.g., EPA.

[00112] In another embodiment, a composition (e.g., a pharmaceutical and/or immunogenic composition) comprises at least the *E. coli* O1A, O2, glucosylated O4, O6A and O25B antigen polysaccharides, preferably bioconjugates of the O1A, O2, glucosylated O4, O6A and O25B

antigen polysaccharides covalently linked to a carrier protein, e.g., EPA (i.e., a pentavalent composition).

[00113] In a preferred embodiment, a composition (e.g., a pharmaceutical and/or immunogenic composition) comprises at least the *E. coli* O1A, O2, glucosylated O4, O6A, O8, O15, O16, O25B and O75 antigen polysaccharides, preferably bioconjugates of the O1A, O2, glucosylated O4, O6A, O8, O15, O16, O25B and O75 antigen polysaccharides covalently linked to a carrier protein, e.g., EPA (i.e., a 9-valent composition).

[00114] In another preferred embodiment, a composition (e.g., a pharmaceutical and/or immunogenic composition) comprises at least the *E. coli* O1A, O2, glucosylated O4, O6A, O8, O15, O16, O18A, O25B and O75 antigen polysaccharides, preferably bioconjugates of the O1A, O2, glucosylated O4, O6A, O8, O15, O16, O18A, O25B and O75 antigen polysaccharides covalently linked to a carrier protein, e.g., EPA (i.e., a 10-valent composition).

[00115] Also contemplated herein are compositions which optionally further comprise additional O-antigens (e.g., in isolated form, or as part of a glycoconjugate or bioconjugate) from other *E. coli* serotypes.

[00116] In some embodiments, each of the additional *E. coli* O1A, O2, O6A, O8, O15, O16, O18A, O25B, and/or O75 antigen polysaccharides is covalently linked to a carrier protein. The O-antigen polysaccharide can be linked to a carrier protein by chemical or other synthetic methods, or the O-antigen polysaccharide can be part of a bioconjugate, and is preferably part of a bioconjugate. Any carrier protein known to those skilled in the art in view of the present disclosure can be used. Suitable carrier proteins include, but are not limited to, detoxified Exotoxin A of *P. aeruginosa* (EPA), *E. coli* flagellin (FliC), CRM197, maltose binding protein (MBP), Diphtheria toxoid, Tetanus toxoid, detoxified hemolysin A of *S. aureus*, clumping factor A, clumping factor B, *E. coli* heat labile enterotoxin, detoxified variants of *E. coli* heat labile enterotoxin, Cholera toxin B subunit (CTB), cholera toxin, detoxified variants of cholera toxin, *E. coli* Sat protein, the passenger domain of *E. coli* Sat protein, *Streptococcus pneumoniae* Pneumolysin, Keyhole limpet hemocyanin (KLH), *P. aeruginosa* PcrV, outer membrane protein of *Neisseria meningitidis* (OMPC), and protein D from non-typeable *Haemophilus influenzae*. Preferably, the carrier protein is EPA.

[00117] In some embodiments, each of the additional *E. coli* O1A, O2, O6A, O8, O15, O16, O18A, O25B, and/or O75 antigen polysaccharides, particularly when part of a bioconjugate, is covalently linked to an asparagine (Asn) residue in the carrier protein, wherein the Asn residue is present in a glycosylation site comprising a glycosylation consensus sequence Asn-X-Ser(Thr), wherein X can be any amino acid except Pro (SEQ ID NO: 1), preferably wherein the Asn residue is present in a glycosylation site comprising a glycosylation consensus sequence Asp(Glu)-X-Asn-Z-Ser(Thr), wherein X and Z are independently selected from any amino acid except Pro (SEQ ID NO: 2). The carrier protein can comprise 1-10 glycosylation sites, preferably 2 to 4 glycosylation sites, most preferably 4 glycosylation sites, each comprising a glycosylation consensus sequence. In a particular embodiment, the carrier protein is EPA-4 carrier protein, for instance EPA-4 carrier protein comprising the amino acid sequence of SEQ ID NO: 3.

[00118] In a particular embodiment, provided herein is a composition (e.g., pharmaceutical and/or immunogenic composition) comprising: (i) a bioconjugate of an *E. coli* glucosylated O4 antigen polysaccharide covalently linked to a detoxified Exotoxin A of *P. aeruginosa* carrier protein comprising SEQ ID NO: 3 (EPA-4 carrier protein), wherein the *E. coli* glucosylated O4 antigen polysaccharide comprises the structure of Formula (O4-Glc+); (ii) a bioconjugate of an *E. coli* O1A antigen polysaccharide covalently linked to an EPA-4 carrier protein, wherein the *E. coli* O1A antigen polysaccharide comprises the structure of Formula (O1A); (iii) a bioconjugate of an *E. coli* O2 antigen polysaccharide covalently linked to an EPA-4 carrier protein, wherein the *E. coli* O2 antigen polysaccharide comprises the structure of Formula (O2); (iv) a bioconjugate of an *E. coli* O6A antigen polysaccharide covalently linked to an EPA-4 carrier protein, wherein the *E. coli* O6A antigen polysaccharide comprises the structure of Formula (O6A); (v) a bioconjugate of an *E. coli* O8 antigen polysaccharide covalently linked to an EPA-4 carrier protein, wherein the *E. coli* O8 antigen polysaccharide comprises the structure of Formula (O8); (vi) a bioconjugate of an *E. coli* O15 antigen polysaccharide covalently linked to an EPA-4 carrier protein, wherein the *E. coli* O15 antigen polysaccharide comprises the structure of Formula (O15); (vii) a bioconjugate of an *E. coli* O16 antigen polysaccharide covalently linked to an EPA-4 carrier protein, wherein the *E. coli* O16 antigen polysaccharide comprises the structure of Formula (O16); (viii) a bioconjugate of an *E. coli* O25B antigen polysaccharide

covalently linked to an EPA-4 carrier protein, wherein the *E. coli* O25B antigen polysaccharide comprises the structure of Formula (O25B); and (ix) a bioconjugate of an *E. coli* O75 antigen polysaccharide covalently linked to an EPA-4 carrier protein, wherein the *E. coli* O75 antigen polysaccharide comprises the structure of Formula (O75), wherein each of the Formulas is provided in Table 1, and for each of the Formulas independently n is an integer of 1 to 100, e.g. 1 to 50, preferably 3 to 50, e.g. 5 to 40.

[00119] In a particular embodiment, said composition (e.g. pharmaceutical and/or immunogenic composition) further comprises: (x) a bioconjugate of an *E. coli* O18A antigen polysaccharide covalently linked to an EPA-4 carrier protein, wherein the *E. coli* O18A antigen polysaccharide comprises the structure of Formula (O18A) as shown in Table 1, wherein n for this structure is is an integer of 1 to 100, e.g. 1 to 50, preferably 3 to 50, e.g. 5 to 40.

[00120] In some embodiments, a composition provided herein comprises a biconjugate of an *E. coli* glucosylated O4 antigen polysaccharide, and at least a bioconjugate of an *E. coli* O25B antigen polysaccharide, wherein the bioconjugate of the *E. coli* O25B antigen polysaccharide is present in the composition at a concentration that is about 1.5 to 6 times, e.g. about 2 to 4 times higher, such as 1.5, 2, 3, 4, 5 or 6 times higher than the concentration of any of the other bioconjugates present in the composition.

[00121] In particular embodiments, a composition comprises bioconjugates of *E. coli* O1A, O2, glucosylated O4, O6A, O8, O15, O16, O25B, and O75 antigen polysaccharides, wherein the bioconjugates of O1A:O2:glucosylated O4:O6A:O8:O15:O16:O25B:O75 are present in a ratio (by weight of O-antigen polysaccharide) of 1:1:1:1:1:1:2:1, or 2:1:1:2:1:1:1:4:1.

[00122] In particular embodiments, a composition comprises bioconjugates of *E. coli* O1A, O2, glucosylated O4, O6A, O8, O15, O16, O18A, O25B, and O75 antigen polysaccharides, wherein the bioconjugates of O1A:O2:glucosylated O4:O6A:O8:O15:O16:O18A:O25B:O75 are present in a ratio (by weight of O-antigen polysaccharide) of 1:1:1:1:1:1:1:2:1, or 2:1:1:2:1:1:1:4:1.

[00123] In some embodiments, a composition provided herein comprises a bioconjugate of an *E. coli* glucosylated O4 antigen polysaccharide, and at least a bioconjugate of an *E. coli* O25B antigen polysaccharide, wherein the bioconjugate of the *E. coli* O25B antigen polysaccharide is present in the composition at a concentration of 2 to 50 µg/mL, preferably 8 to 40 µg/mL, more

preferably 16-32 µg/mL, such as 16, 18, 20, 22, 24, 26, 28, 30, or 32 µg/mL. In such embodiments, the concentration of the bioconjugate of the *E. coli* O25B antigen polysaccharide is preferably about 1.5 to 6 times, e.g., about 2 to 4 times higher, such as 1.5, 2, 3, 4, 5, or 6 times higher than the concentration of any of the other bioconjugates present in the composition.

[00124] In certain embodiments, the compositions described herein (e.g., pharmaceutical and/or immunogenic compositions) comprise, or are administered in combination with, an adjuvant. The adjuvant for administration in combination with a composition described herein may be administered before (e.g. within 72 hours, 48 hours, 24 hours, 12 hours, 6 hours, 2 hours, 1 hour, 10 minutes), concomitantly with, or after (e.g. within 72 hours, 48 hours, 24 hours, 12 hours, 6 hours, 2 hours, 1 hour, 10 minutes) administration of said composition. As used herein, the term “adjuvant” refers to a compound that when administered in conjunction with or as part of a composition described herein augments, enhances and/or boosts the immune response to an *E. coli* O-antigen polysaccharide in a bioconjugate, but when the adjuvant compound is administered alone does not generate an immune response to the *E. coli* O-antigen polysaccharide in the bioconjugate. In some embodiments, the adjuvant enhances an immune response to an *E. coli* O-antigen polysaccharide in a bioconjugate thereof and does not produce an allergy or other adverse reaction. Adjuvants can enhance an immune response by several mechanisms including, e.g., lymphocyte recruitment, stimulation of B and/or T cells, and stimulation of macrophages.

[00125] Examples of suitable adjuvants include, but are not limited to, aluminum salts (alum) (such as aluminum hydroxide, aluminum phosphate, aluminum sulfate and aluminum oxide, including nanoparticles comprising alum or nanoalum formulations), calcium phosphate, monophosphoryl lipid A (MPL) or 3-de-O-acylated monophosphoryl lipid A (3D-MPL) (see e.g., United Kingdom Patent GB2220211, EP0971739, EP1194166, US6491919), AS01, AS02, AS03 and AS04 (all GlaxoSmithKline; see e.g. EP1126876, US7357936 for AS04, EP0671948, EP0761231, US5750110 for AS02), MF59 (Novartis), imidazopyridine compounds (see WO2007/109812), imidazoquinoxaline compounds (see WO2007/109813), delta-inulin, STING-activating synthetic cyclic-di-nucleotides (e.g. US20150056224), combinations of lecithin and carbomer homopolymers (e.g. US6676958), and saponins, such as QuilA and QS21 (see e.g. Zhu D and W Tuo, 2016, Nat Prod Chem Res 3: e113 (doi:10.4172/2329-6836.1000e113), Matrix M,

Iscoms, Iscomatrix, etc, optionally in combination with QS7 (see Kensil *et al.*, in Vaccine Design: The Subunit and Adjuvant Approach (eds. Powell & Newman, Plenum Press, NY, 1995); U.S. Pat. No. 5,057,540). In some embodiments, the adjuvant is Freund's adjuvant (complete or incomplete). Other adjuvants are oil in water emulsions (such as squalene or peanut oil), optionally in combination with immune stimulants, such as monophosphoryl lipid A (see Stoute *et al.*, N. Engl. J. Med. 336, 86-91 (1997)). Another adjuvant is CpG (Bioworld Today, Nov. 15, 1998). Further examples of adjuvants are liposomes containing immune stimulants such as MPL and QS21 such as in AS01E and AS01B (e.g. US 2011/0206758). Other examples of adjuvants are CpG (Bioworld Today, Nov. 15, 1998) and imidazoquinolines (such as imiquimod and R848). See, e.g., Reed G, et al., 2013, *Nature Med.*, 19: 1597-1608. In certain embodiments, the adjuvant contains a toll-like receptor 4 (TLR4) agonist. TLR4 agonists are well known in the art, see e.g. Ireton GC and SG Reed, 2013, *Expert Rev Vaccines* 12: 793-807. In certain embodiments, the adjuvant comprises a TLR4 agonist comprising lipid A, or an analog or derivative thereof, such as MPL, 3D-MPL, RC529 (e.g. EP1385541), PET-lipid A, GLA (glycopyranosyl lipid adjuvant, a synthetic disaccharide glycolipid; e.g. US20100310602, US8722064), SLA (e.g. Carter D et al, 2016, *Clin Transl Immunology* 5: e108 (doi: 10.1038/cti.2016.63), which describes a structure-function approach to optimize TLR4 ligands for human vaccines), PHAD (phosphorylated hexaacyl disaccharide), 3D-PHAD (the structure of which is the same as that of GLA), 3D-(6-acyl)-PHAD (3D(6A)-PHAD) (PHAD, 3D-PHAD, and 3D(6A)PHAD are synthetic lipid A variants, see e.g. avantilipids.com/divisions/adjuvants, which also provide structures of these molecules), E6020 (CAS Number 287180-63-6), ONO4007, OM-174, and the like.

[00126] In certain embodiments, the compositions described herein do not comprise, and are not administered in combination with, an adjuvant.

[00127] In certain embodiments, the compositions described herein are formulated to be suitable for the intended route of administration to a subject. For example, the compositions (e.g., pharmaceutical and/or immunogenic) described herein can be formulated for subcutaneous, parenteral, oral, sublingual, buccal, intradermal, transdermal, colorectal, intraperitoneal, rectal administration, intravenous, intranasal, intratracheal, intramuscular, topical, transdermal, or

intradermal administration. In a specific embodiment, a composition provided herein (e.g., pharmaceutical and/or immunogenic) is formulated for intramuscular injection.

[00128] Methods of Use

[00129] Bioconjugates and compositions provided herein can be used to induce antibodies against an *E. coli* glucosylated O4 antigen in a subject, and to vaccinate a subject against *E. coli* in particular extra-intestinal pathogenic *E. coli* (ExPEC). As used herein, “subject” means any animal, preferably a mammal, to whom will be or has been administered a bioconjugate or composition provided herein. The term “mammal” as used herein, encompasses any mammal. Examples of mammals include, but are not limited to, cows, horses, sheep, pigs, cats, dogs, mice, rats, rabbits, guinea pigs, non-human primates (NHPs) such as monkeys or apes, humans, etc. In certain embodiments, a subject is a human. A human subject may be of any age. In certain embodiments, a subject is a human of about two months to about 18 years old, e.g. of 1 year to 18 years old. In certain embodiments, a subject is a human of at least 18 years old. In certain embodiments, a subject is a human of 15 to 50 years old, e.g. 18 to 45 years old, e.g. 20 to 40 years old. In certain embodiments, a subject is a human male. In certain embodiments, a subject is a human female. In certain embodiments, a subject is immunocompromised. In certain embodiments, a subject is a human of at least 50 years, at least 55 years, at least 60 years, at least 65 years old. In certain embodiments, a subject is a human that is not older than 100 years, not older than 95 years, not older than 90 years, not older than 85 years, not older than 80 years, or not older than 75 years. In certain embodiments, a subject is a human of at least 60 years old, and not older than 85 years old. In certain embodiments, a subject is a human in stable health. In certain embodiments, a subject is a human adult of at least 60 and not more than 85 years old in stable health. In certain embodiments, a subject is a human that has a history of a urinary tract infection (UTI, i.e. a bacterial infection in the urethra, bladder, ureters, and/or kidneys), i.e. having had at least one UTI episode in his or her life. In certain embodiments, a subject is a human that has a history of UTI in the past twenty, fifteen, twelve, ten, nine, eight, seven, six, five, four, three, two or one years. In certain embodiments, a subject is a human that has a history of UTI in the past two years. In certain embodiments, a subject is a human subject that has a history of recurrent UTI, i.e. having had at least two UTIs in six months or at least three UTIs in one year. In certain embodiments, a subject is a human subject that has a history of recurrent UTI

in the past two years. In certain embodiments, a subject is a human of 60 years or older in stable health. In certain embodiments, a subject is a human of 60 years or older that has a history of UTI in the past two years. In certain embodiments, a subject is a human of at least 60 years and less than 75 years old that has a history of UTI in the past two years. In certain embodiments, a subject is a human subject of 75 years or older that has a history of UTI in the past two years. In certain embodiments, a subject is a patient scheduled for undergoing elective urogenital and/or abdominal procedures or surgeries, e.g. transrectal ultrasound-guided prostate needle biopsy (TRUS-PNB).

[00130] In one aspect, provided herein is a method of inducing antibodies against an *E. coli* glucosylated O4 antigen in a subject, comprising administering to the subject any of the bioconjugates of an *E. coli* glucosylated O4 antigen covalently linked to a carrier protein described herein, or a composition comprising a bioconjugate of an *E. coli* glucosylated O4 antigen covalently linked to a protein, alone or further in combination with other *E. coli* O-antigen polysaccharides or glycoconjugates or bioconjugates thereof.

[00131] In certain embodiments, the antibodies induced, elicited or identified against an *E. coli* glucosylated O4 antigen have opsonophagocytic activity. In particular embodiments, the antibodies induced, elicited or identified are cross-reactive antibodies capable of mediating opsonophagocytic killing of both *E. coli* glucosylated and non-glucosylated O4 strains.

[00132] In certain embodiments, the antibodies induced, elicited or identified identified against an *E. coli* glucosylated O4 antigen specifically recognize unmodified and glucose modified O4 antigen polysaccharide. In certain embodiments, the antibodies induced, elicited or identified against an *E. coli* glucosylated O4 antigen specifically recognize *E. coli* of the O4 serotype. In certain embodiments, the antibodies induced by a bioconjugate of an *E. coli* glucosylated O4 antigen bind preferentially to glucosylated O4 antigen as compared to non-glucosylated O4 antigen.

[00133] Antibodies induced by the bioconjugates and compositions described herein can include immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site that specifically binds to an *E. coli* O-antigen polysaccharide, e.g., glucosylated O4 antigen polysaccharide.

[00134] Antibodies induced, elicited or identified using the bioconjugates or compositions provided herein can be used to monitor the efficacy of a therapy and/or disease progression. Any immunoassay system known in the art can be used for this purpose including, but not limited to, competitive and noncompetitive assay systems using techniques such as radioimmunoassays, ELISA (enzyme linked immunosorbent assays), electrochemiluminescence (ECL)-based immunoassays, “sandwich” immunoassays, precipitin reactions, gel diffusion precipitin reactions, immunodiffusion assays, immunoradiometric assays, fluorescent immunoassays, protein A immunoassays and immunoelectrophoresis assays. Several of these assays, e.g. ECL-based immunoassays, can be done in multiplex format, and typically multiplex assay formats are preferred.

[00135] Antibodies induced, elicited or identified using a bioconjugate of an *E. coli* glucosylated O4 antigen polysaccharide can be used to detect *E. coli* O4 strains, particularly glucosylated O4 strains, for example, from a plurality of *E. coli* strains and/or to diagnose an infection by an *E. coli* O4 or glucosylated O4 strain.

[00136] In another aspect, provided herein is a method of vaccinating a subject against *E. coli* (e.g. extra-intestinal pathogenic *E. coli*, ExPEC), comprising administering to the subject any of the bioconjugates of an *E. coli* glucosylated O4 antigen covalently linked to a carrier protein described herein, or a composition comprising a bioconjugate of an *E. coli* glucosylated O4 antigen covalent linked to a carrier protein, alone or further in combination with other *E. coli* O-antigens or glycoconjugates or bioconjugates thereof. One skilled in the art will understand that the subject will be vaccinated against *E. coli* strains whose O antigens or glycoconjugates or bioconjugates thereof are present in the composition administered. For example, administration of a composition comprising O1A, O2, glucosylated O4, O6A, and O25B antigen polysaccharides can be used to a vaccinate a subject against *E. coli* serotypes O1A, O2, O4, O6A, and O25B.

[00137] In certain embodiments, vaccination is for preventing an invasive ExPEC disease (IED), e.g., urosepsis, bacteremia, sepsis, etc. In certain embodiments, vaccination is to prevent or reduce the occurrence or severity of urinary tract infections. In certain embodiments, an IED can be hospital-acquired, e.g. in patients undergoing urogenital and/or abdominal procedures or surgeries. In certain embodiments, an IED can be healthcare-associated, e.g. in patients receiving

health care for another condition, for instance via central lines, catheters, etc, e.g. in a hospital, ambulatory surgical center end-stage renal disease facility, long-term care facility, etc. In certain embodiments, the IED can be community-acquired, e.g. in a patient that was not recently exposed to healthcare risks.

[00138] In another aspect, provided herein is a method of inducing an immune response against *E. coli* (e.g., ExPEC) in a subject, comprising administering to the subject any of the bioconjugates of an *E. coli* glucosylated O4 antigen covalently linked to a carrier protein described herein, or a composition comprising a bioconjugate of an *E. coli* glucosylated O4 antigen covalently linked to a carrier protein, alone or further in combination with other *E. coli* O-antigens or glycoconjugates or bioconjugates thereof. In one embodiment, the subject has an *E. coli* (e.g., ExPEC) infection at the time of administration. In a preferred embodiment, the subject does not have an *E. coli* (e.g., ExPEC) infection at the time of administration.

[00139] In certain embodiments, the compositions and bioconjugates described herein can be administered to a subject to induce an immune response that includes the production of antibodies, preferably antibodies having opsonophagocytic activity. Such antibodies can be isolated using techniques known to one of skill in the art (e.g., immunoaffinity chromatography, centrifugation, precipitation, etc.).

[00140] The ability of the bioconjugates and compositions described herein to generate an immune response in a subject can be assessed using any approach known to those of skill in the art or described herein. In some embodiments, the ability of a bioconjugate to generate an immune response in a subject can be assessed by immunizing a subject (e.g., a mouse, rat, rabbit, or monkey) or set of subjects with a bioconjugate described herein and immunizing an additional subject (e.g., a mouse, rat, rabbit, or monkey) or set of subjects with a control (PBS). The subjects or set of subjects can subsequently be challenged with ExPEC and the ability of the ExPEC to cause disease (e.g., UTI, bacteremia, or other disease) in the subjects or set of subjects can be determined. Those skilled in the art will recognize that if the subject or set of subjects immunized with the control suffer(s) from disease subsequent to challenge with the ExPEC but the subject or set of subjects immunized with a bioconjugate(s) or composition thereof described herein suffer less from or do not suffer from disease, then the bioconjugate is able to generate an immune response in a subject. The ability of a bioconjugate(s) or composition thereof described

herein to induce antiserum that cross-reacts with an O antigen from ExPEC can be tested by, e.g., an immunoassay, such as an ELISA (see e.g., Van den Doppelsteen et al, 2016, Vaccine 34: 4152-4160), or an ECL-based immunoassay.

[00141] For example, the ability of the bioconjugates described herein to generate an immune response in a subject can be assessed using a serum bactericidal assay (SBA) or opsonophagocytic killing assay (OPK assay, or OPKA), which represents an established and accepted method that has been used to obtain approval of glycoconjugate-based vaccines. Such assays are well-known in the art and, briefly, comprise the steps of generating and isolating antibodies against a target of interest (e.g., an O antigen polysaccharide, e.g., *E. coli* glucosylated O4 antigen polysaccharide) by administering to a subject (e.g., a mouse, rat, rabbit, or monkey) a compound that elicits such antibodies. Subsequently, the bactericidal capacity of the antibodies can be assessed by, e.g., culturing the bacteria in question (e.g., *E. coli* of the relevant serotype) in the presence of the antibodies and complement and – depending on the assay - neutrophilic cells and assaying the ability of the antibodies to mediate killing and/or neutralization of the bacteria, e.g., using standard microbiological approaches. For an example of OPK assay for *E.coli* bioconjugate vaccines, see e.g. Abbanat et al, 2017, Clin. Vaccine Immunol. 24: e00123-17. An OPK assay can be performed in monoplex or multiplex format, of which multiplex format (e.g. testing multiple serotypes at the same time) is typically preferred. A multiplex OPK assay is sometimes referred to herein as ‘MOPA’.

[00142] In some embodiments, the methods described herein comprise administering an effective amount of bioconjugates of an *E. coli* glucosylated O4 antigen covalently linked to a carrier protein described herein, or a composition comprising a bioconjugate of an *E. coli* glucosylated O4 antigen covalently linked to a carrier protein, alone or further in combination with other *E. coli* O-antigens or glycoconjugates or bioconjugates thereof. In one embodiment, an “effective amount” is an amount that vaccinates a subject against *E. coli* (e.g., ExPEC). In another embodiment, an “effective amount” is an amount that induces an immune response against *E. coli* (e.g., ExPEC) in a subject, such as an immune response including the production of antibodies, preferably antibodies having opsonophagocytic activity.

[00143] In particular embodiments, wherein a composition provided herein comprises a bioconjugate of an *E. coli* glucosylated O4 antigen polysaccharide and at least a bioconjugate of

an *E. coli* O25B antigen polysaccharide, an effective amount of the *E. coli* O25B antigen polysaccharide is about 1.5 to 6 times, e.g. about 2 to 4 times higher, such as 1.5, 2, 3, 4, 5 or 6 times higher than the concentration of any of the other bioconjugates present in the composition. In such embodiments, an effective amount of the *E. coli* O25B antigen polysaccharide is for instance about 5 to 18 µg per administration, such as 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 µg per administration.

[00144]

[00145] In certain embodiments, a bioconjugate or composition according to the invention is administered to a subject once. In certain embodiments, a bioconjugate or composition according to the invention is administered to a subject more than once, e.g. in a prime-boost regimen. In certain embodiments, the time between two administrations is at least two weeks, at least one month, at least two months, at least three months, at least six months, at least one year, at least two years, at least five years, at least ten years, or at least fifteen years. In humans, a desired immune response can typically be generated by a single administration of a bioconjugate or composition according to the invention. In certain embodiments, a repeat administration after for instance ten years is provided.

[00146] Host Cells

[00147] Provided herein are host cells, e.g., prokaryotic host cells, capable of producing *E. coli* O antigens and bioconjugates comprising such *E. coli* O antigens. The host cells provided herein preferably are modified to comprise (e.g., through genetic engineering) one or more of the nucleic acids encoding host cell machinery (e.g., glycosyltransferases) used to produce *E. coli* O-antigen polysaccharides and/or bioconjugates thereof.

[00148] Any host cells known to those of skill in the art can be used to produce the *E. coli* O antigen polysaccharides described herein (e.g., *E. coli* glucosylated O4 antigen polysaccharide) and bioconjugates comprising the *E. coli* O antigen polysaccharides described herein (e.g., a bioconjugate of *E. coli* glucosylated O4 antigen polysaccharide) including archaea, prokaryotic host cells, and eukaryotic host cells. In a preferred embodiment, a host cell is a prokaryotic host cell. Exemplary prokaryotic host cells for use in production of the *E. coli* O antigen polysaccharides described herein and bioconjugates comprising the *E. coli* O antigen polysaccharides described herein include, but are not limited to, *Escherichia* species, *Shigella*

species, *Klebsiella* species, *Xhantomonas* species, *Salmonella* species, *Yersinia* species, *Lactococcus* species, *Lactobacillus* species, *Pseudomonas* species, *Corynebacterium* species, *Streptomyces* species, *Streptococcus* species, *Staphylococcus* species, *Bacillus* species, and *Clostridium* species.

[00149] In a specific embodiment, the host cell used to produce the *E. coli* O antigen polysaccharides described herein and bioconjugates comprising the *E. coli* O antigen polysaccharides described herein is a prokaryotic host cell, and is preferably *E. coli*.

[00150] In certain embodiments, the host cells used to produce the *E. coli* O antigen polysaccharides and bioconjugates described herein are engineered to comprise heterologous nucleic acids, e.g., heterologous nucleic acids comprising *rfb* gene clusters of a desired O antigen serotype, heterologous nucleic acids that encode one or more carrier proteins and/or glycosyltransferases. In a specific embodiment, heterologous *rfb* genes, and/or heterologous nucleic acids that encode proteins involved in glycosylation pathways (e.g., prokaryotic and/or eukaryotic glycosylation pathways) can be introduced into the host cells described herein. Such nucleic acids can encode proteins including, but not limited to, oligosaccharyl transferases and/or glycosyltransferases.

[00151] Sequences of various genes and gene clusters encoding glycosyltransferases useful in making recombinant host cells that can, e.g., be used to prepare *E. coli* O antigen polysaccharides and bioconjugates thereof are described herein. Those skilled in the art will appreciate that due to the degeneracy of the genetic code, a protein having a specific amino acid sequence can be encoded by multiple different nucleic acids. Thus, those skilled in the art will understand that a nucleic acid provided herein can be altered in such a way that its sequence differs from a sequence provided herein, without affecting the amino acid sequence of the protein encoded by the nucleic acid.

[00152] Provided herein are host cells (e.g., recombinant host cells) for producing a bioconjugate of an *E. coli* glucosylated O4 antigen polysaccharide, O1A antigen polysaccharide, O2 antigen polysaccharide, O6A antigen polysaccharide, O8 antigen polysaccharide, O15 antigen polysaccharide, O16 antigen polysaccharide, O18A antigen polysaccharide, O25B antigen polysaccharide, or O75 antigen polysaccharide. The host cells provided herein comprise nucleic acids encoding enzymes (e.g., glycosyltransferases) capable of producing the *E. coli* O

antigen polysaccharide. The host cells provided herein can naturally express nucleic acids capable of producing an O antigen of interest, or the host cells can be made to express such nucleic acids. In certain embodiments the nucleic acids are heterologous to the host cells and introduced into the host cells using genetic approaches known in the art. For example, the nucleic acids can be introduced into the host cell by genetic manipulation (e.g., the gene cluster is expressed on a plasmid or plasmids or integrated into the host cell genome (see, e.g., International Patent Application Publications WO 2014/037585, WO 2014/057109, WO 2015/052344).

[00153] In one embodiment, provided herein is a host cell (e.g., recombinant host cell) capable of producing a bioconjugate of an *E. coli* glucosylated O4 antigen polysaccharide covalently linked to a carrier protein. Such a host cell comprises, preferably by engineering a precursor cell, a nucleic acid sequence encoding a *gtrS* gene, which, to the best of the knowledge of the inventors, was identified herein for the first time as encoding a polysaccharide branching enzyme capable of transferring glucose to the *E. coli* O4 antigen (i.e., a glucosyltransferase specific to the *E. coli* O4 antigen polysaccharide), and particularly to L-Rha via an α -1,3-glycosidic linkage. An example of an amino acid sequence of such branching enzyme is provided in SEQ ID NO: 4. Other examples comprise amino acid sequences that are at least 80% identical thereto.

Exemplary examples of nucleic acid sequence encoding *gtrS* genes specific to the *E. coli* O4 antigen polysaccharide include, but are not limited to, SEQ ID NO: 5, or degenerate nucleic acid sequences thereto that encode SEQ ID NO: 4, or nucleic acid sequences that encode functional O4-specific GtrS enzymes that have at least 80% identity to SEQ ID NO: 4.

[00154] In a specific embodiment, a host cell (e.g., recombinant host cell) capable of producing a bioconjugate of an *E. coli* glucosylated O4 antigen polysaccharide covalently linked to a carrier protein, comprises a nucleotide sequence encoding a glucosyl transferase having at least 80% sequence identity to SEQ ID NO: 4, such as about 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 95%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 4. In view of the redundancy in the genetic code, one of ordinary skill in the art can make variants of nucleic acid sequences encoding the amino acid sequences of glucosyl transferases, e.g., using codon optimized sequences, if desired.

[00155] In certain embodiments, a host cell (e.g., recombinant host cell) capable of producing a bioconjugate of an *E. coli* glucosylated O4 antigen polysaccharide covalently linked to a carrier protein, comprising a nucleotide sequence encoding a glucosyl transferase (GtrS) having at least 80% sequence identity to SEQ ID NO: 4, further comprises a nucleotide sequence encoding a bactoprenol-linked glucose translocase (GtrA) having at least 80% sequence identity to SEQ ID NO: 7, and a nucleotide sequence encoding a bactoprenol glucosyl transferase (GtrB) having at least 80% sequence identity to SEQ ID NO: 8. In certain embodiments, said nucleic acid sequences encode GtrA and GtrB proteins that are at least about 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NOs: 7 and 8, respectively, and have bactoprenol-linked glucose translocase (SEQ ID NO: 7) and bactoprenol glucosyl transferase (SEQ ID NO: 8) activity, respectively. In view of the redundancy in the genetic code, one of ordinary skill in the art can make variants of nucleic encoding the amino acid sequences of bactoprenol-linked glucose translocases and of bactoprenol glucosyl transferases, e.g., using codon optimized sequences, if desired.

[00156] A host cell (e.g., recombinant host cell) capable of producing a bioconjugate of an *E. coli* glucosylated O4 antigen polysaccharide covalently linked to a carrier protein provided herein further comprises a nucleotide sequence of an *rfb* gene cluster for the *E. coli* O4 antigen polysaccharide. An example of an *rfb* gene cluster useful for production of the *E. coli* O4 antigen polysaccharide is provided herein as SEQ ID NO: 9. Another example can be found in GenBank, locus AY568960. Degenerate nucleic acid sequences encoding the same enzymes as encoded by this sequence, or sequences that encode enzymes that are at least 80% identical, preferably at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% identical, can also be used.

[00157] In a specific embodiment, provided herein is a host cell (e.g., a recombinant host cell, preferably a recombinant prokaryotic host cell, preferably a recombinant *E. coli* host cell) that produces glucosylated O4 antigen polysaccharide, wherein the host cell comprises *gtrS*, an *rfb* gene cluster for the *E. coli* O4 antigen polysaccharide, and nucleic acid encoding a carrier protein. Such host cells can be engineered using recombinant approaches to comprise one or more plasmids comprising the *gtrS* gene, the *rfb* gene cluster, and/or nucleic acid encoding a carrier protein, or to comprise some or all of the relevant genes such as *gtrS*, the *rfb* cluster

and/or the nucleic acid encoding the carrier protein integrated into the host cell genome. In certain embodiments, the genes or gene clusters have been integrated into the genome of the host cell using homologous recombination. An advantage of integration of genes into the genome of the host cell is stability in the absence of antibiotic selection.

[00158] In another specific embodiment, provided herein is a host cell (e.g., a recombinant host cell, preferably a recombinant prokaryotic host cell) that produces glucosylated O4 antigen polysaccharide, wherein the host cell comprises GtrS (glucosyltransferase), as well as the enzymes encoded by the O4 *rfb* cluster. In certain embodiments, some or all of the aforementioned enzymes are heterologous to the host cell.

[00159] In other specific embodiments, provided herein is a host cell (e.g. a recombinant host cell, preferably a recombinant prokaryotic host cell) that produces *E. coli* glucosylated O4 antigen polysaccharide, preferably a bioconjugate of *E. coli* glucosylated O4 antigen polysaccharide, wherein the host cell further comprises a nucleotide sequence encoding an oligosaccharyl transferase and/or a nucleotide sequence encoding a carrier protein. In one specific embodiment, the oligosaccharyl transferase is heterologous to the host cell. In another specific embodiment, the carrier protein is heterologous to the host cell. Preferably, the host cell comprises a heterologous nucleotide sequence encoding a glucosyl transferase having at least 80% sequence identity to SEQ ID NO: 4. In preferred embodiments, the *rfb* genes of the O4 cluster are heterologous to the host cell. Preferably the sequence encoding the enzyme that is capable of introducing the branched glucose side chain to the O4 antigen, i.e. the *gtrS* gene (encoding a glucosyl transferase having at least 80% sequence identity to SEQ ID NO:4) is heterologous to the host cell. A nucleic acid is heterologous to the host cell if the same sequence is not naturally present in said host cell. Heterologous nucleic acid can for instance be introduced in a parent cell by genetic engineering, e.g by transformation (e.g. chemical transformation or electroporation) and/or recombination. In certain embodiments, heterologous nucleic acid such as a desired *rfb* locus, *gtrS* coding sequence, carrier protein encoding sequence, and/or glycosyltransferase encoding sequence are integrated into the genome of the host cell, preferably a bacterial host cell, preferably an *E. coli* host cell. In preferred embodiments, the endogenous *rfb* locus and if applicable *gtrS* coding sequence have been inactivated, preferably deleted from the genome of the recombinant host cell as compared to a predecessor thereof, and preferably

these are replaced by the desired heterologous *rfb* locus, and if applicable desired *gtrS* coding sequence, respectively. In certain embodiments the host cell is a K-12 of *E. coli* (as a non-limiting example, *E. coli* strain W3110 is a K-12 strain), or a B strain of *E. coli* (as a non-limiting example, *E. coli* strain BL21 is a B strain), or any other well-defined strain of *E. coli*, e.g. laboratory strains or production strains, in contrast to primary wild-type isolates. In preferred embodiments, the host cell is derived from *E. coli* that does not express O4 antigen or glucosylated O4 antigen, by introduction into such *E. coli* of the O4 *rfb* locus and the *gtrS* gene encoding a glucosyl transferase having at least 80% sequence identity to SEQ ID NO:4.

Advantages of using well-characterized strains, such as *E. coli* K-12 or *E. coli* B, as precursors for host cells is the possibility to use a similar production process for different O-antigen bioconjugates, since the characteristics of the production strain are well-defined. Even though bioconjugates of different O-antigens will behave differently and expression processes can be optimized per production strain, at least the basic process for production of O-antigen bioconjugates will be more predictable using such well-defined precursor strains than when unknown strains such as wild-type isolates are used as precursors for production of host strains. This way, experience with production of earlier described *E. coli* O-antigen bioconjugates such as O1A, O2, O6A and O25B bioconjugates as described in for instance WO 2015/124769 and WO 2017/035181 can be used as basis to design production of other *E. coli* O-antigen bioconjugates. Unlike *gtrS*, the *gtrA* and *gtrB* genes are not serotype-specific, and in certain embodiments these are homologous to the host cell (e.g. *E. coli* K12 strain W3110 includes *gtrA* and *gtrB* genes that are capable of functioning together with the O4-serotype specific recombinantly introduced *gtrS* gene encoding a glucosyl transferase of SEQ ID NO: 4 or a glucosyl transferase that is at least 80% identical thereto, replacing the endogenous *gtrS* gene). In other embodiments, one or both of *gtrA* and *gtrB* genes (encoding GtrA and GtrB proteins that are at least about 80% identical to SEQ ID NOs: 7 and 8, respectively, and having bactoprenol-linked glucose translocase and bactoprenol glucosyl transferase activity respectively, are also recombinantly introduced in the host cell, for instance in case the host cell does not have endogenous *gtrA* and/or *gtrB* genes.

[00160] Also provided herein are host cells (e.g., recombinant host cells) capable of producing a bioconjugate of an *E. coli* O1A, O2, O6A, O8, O15, O16, O18A, O25B, or O75 antigen

polysaccharide covalently linked to a carrier protein. Such host cells (e.g., recombinant host cells) comprise nucleotide sequence of an *rfb* gene cluster specific to the O-antigen polysaccharide. The *rfb* gene clusters can be isolated from wild-type *E. coli* strains, and combined with nucleic acids encoding an oligosaccharyl transferase (e.g., PglB) and carrier protein (e.g., EPA) within one host cell to obtain a recombinant host cell that produces the *E. coli* O-antigen of interest or bioconjugate thereof. For example, such host cells can be engineered using recombinant approaches to comprise one or more plasmids comprising the *rfb* gene cluster, oligosaccharyl transferase (e.g., PglB) and carrier protein (e.g., EPA) using bioconjugation technology such as that described in WO 2014/037585, WO 2009/104074, and WO 2009/089396. Preferably the host cells comprise the *rfb* gene clusters integrated into their genome. The nucleic acids encoding oligosaccharyl transferase, carrier protein, and where applicable *gtrS* gene, are in certain embodiments also integrated into the genome of the host cell. Heterologous or homologous *gtrA* and *gtrB* genes are in certain embodiments also integrated into the genome of the host cell.

[00161] Preparation of bioconjugates for O1A, O2, O6A and O25B antigens has been described in detail in WO 2015/124769 and WO 2017/035181. Exemplary gene clusters for each *E. coli* O antigen (*rfb* loci) have been described in Iguchi A, et al, DNA Research, 2014, 1-7 (doi: 10.1093/dnarecs/dsu043), and in DebRoy C, et al, PLoS One. 2016, 11(1):e0147434 (doi: 10.1371/journal.pone.0147434; correction in: Plos One. 2016, 11(4):e0154551, doi: 10.1371/journal.pone.0154551). Nucleic acid sequences for the *rfb* clusters and amino acid sequences for proteins encoded therein can also be found in public databases, such as GenBank. Exemplary sequences for *rfb* clusters that can be used in production strains for bioconjugates with polysaccharide antigens of the serotypes disclosed herein, are also provided in SEQ ID NOs: 9 and 11-19. Thus, for each of the desired bioconjugates mentioned above, the respective *rfb* cluster can be introduced into a host cell, to obtain host cells with the specific *rfb* cluster for the desired O-antigen, as well as containing nucleic acid encoding oligosaccharyltransferase and carrier protein. For reasons indicated above, preferably the host cells are recombinant host cells, and preferably are derived from strains with relatively well-known characteristics, such as *E. coli* laboratory or production strains, e.g. *E. coli* K12 or *E. coli* BL21, etc. Preferably, the *rfb* clusters are heterologous to the host cell, e.g. introduced into a precursor cell of the host cell, and

preferably integrated into the genome thereof. Preferably an original *rfb* gene cluster, if such was present in a precursor cell, has been replaced by the *rfb* gene cluster for the O-antigen of interest in the host cell, to enable production of bioconjugate of the O-antigen of interest. Preferably the oligosaccharyltransferase is heterologous to the host cell, and in certain embodiments nucleic acid encoding such oligosaccharyltransferase is integrated into the genome of the host cell.

[00162] Any of the host cells provided herein (e.g., recombinant host cells, preferably recombinant prokaryotic host cells) comprise nucleic acids encoding additional enzymes active in the *N*-glycosylation of proteins, e.g., the host cell provided herein can further comprise a nucleic acid encoding an oligosaccharyl transferase or nucleic acids encoding other glycosyltransferases.

[00163] The host cells provided herein comprise a nucleic acid that encodes an oligosaccharyl transferase. Oligosaccharyl transferases transfer lipid-linked oligosaccharides to asparagine residues of nascent polypeptide chains that comprise an *N*-glycosylation consensus motif. The nucleic acid that encodes an oligosaccharyl transferase can be native to the host cell, or can be introduced into the host cell using genetic approaches. In preferred embodiments, the oligosaccharyl transferase is heterologous to the host cell. *E. coli* does not naturally comprise an oligosaccharyl transferase, and hence if *E.coli* is used as a host cell for production of bioconjugates, a heterologous oligosaccharyl transferase is comprised in such host cell, e.g. upon introduction by genetic engineering. The oligosaccharyl transferase can be from any source known in the art in view of the present disclosure.

[00164] In certain embodiments, an alternative to an oligosaccharyl transferase with *N*-glycosyltransferase activity, such as an *O*-glycosyltransferase, e.g. as a non-limiting example PglL, can be used, in conjunction with its own, different, glycosylation consensus sequence in the carrier protein, as for instance described in WO 2016/82597. Other glycosyltransferases, such as *O*-glycosyltransferases, can thus also be used as an oligosaccharyltransferase according to the invention.

[00165] In certain preferred embodiments, the oligosaccharyl transferase is an oligosaccharyl transferase from *Campylobacter*. For example, in one embodiment, the oligosaccharyl transferase is an oligosaccharyl transferase from *Campylobacter jejuni* (i.e., *pglB*; see, e.g., Wacker et al., 2002, *Science* 298:1790-1793; see also, e.g., NCBI Gene ID: 3231775, UniProt

Accession No. O86154). In another embodiment, the oligosaccharyl transferase is an oligosaccharyl transferase from *Campylobacter lari* (see, e.g., NCBI Gene ID: 7410986).

[00166] In specific embodiments, the oligosaccharyl transferase is PglB oligosaccharyl transferase from *Campylobacter jejuni*, including the natural (wild-type) protein or any variant thereof, such as those described in International Patent Application Publications WO 2016/107818 and WO 2016/107819. PglB can transfer lipid-linked oligosaccharides to asparagine residues in the consensus sequences SEQ ID NO: 1 and SEQ ID NO: 2. In particular embodiments, the PglB oligosaccharyl transferase comprises SEQ ID NO: 6, or a variant thereof. In certain embodiments one or more endogenous glycosylation consensus sequences in a wild-type PglB have been mutated to avoid PglB autoglycosylation, e.g. SEQ ID NO: 6 comprising the mutation N534Q. Examples of variant PglB oligosaccharyl transferases suitable for use in the recombinant host cells provided herein include the PglB oligosaccharyl transferase of SEQ ID NO: 6 comprising at least one mutation selected from the group consisting of N311V, K482R, D483H, A669V, Y77H, S80R, Q287P, and K289R. In one particular embodiment, a variant PglB oligosaccharyl transferase has SEQ ID NO: 6 comprising the mutation N311V. In another particular embodiment, a variant PglB oligosaccharyl transferase has SEQ ID NO: 6 comprising the mutations Y77H and N311V. In another particular embodiment, a variant PglB oligosaccharyl transferase has SEQ ID NO: 6 comprising the mutations N311V, K482R, D483H, and A669V. In another particular embodiment, a variant PglB oligosaccharyl transferase has SEQ ID NO: 6 comprising the mutations Y77H, S80R, Q287P, K289R, and N311V. It was found and described herein that certain PglB oligosaccharyl transferase variants give surprisingly improved yields in production of *E. coli* O-antigen bioconjugates of specific serotypes. The improved or optimal PglB variant for a given *E. coli* O-antigen was not predictable. The invention in certain aspects therefore also provides methods for producing bioconjugates of specific *E. coli* O-antigens, using specific PglB variants as the oligosaccharyl transferase. Further variants of PglB that are at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% identical to SEQ ID NO: 6 and still have oligosaccharyl transferase activity, preferably having one or more of the specific amino acids on the indicated positions disclosed in combination herein (e.g. 77Y, 80S, 287Q, 289K, 311N, 482K, 483D, 669A; or 311V; or 311V,

482R, 483H, 669V; or 77H, 80R, 287P, 289R, 311V; or 77H, 311V; etc) can also be used for production of bioconjugates.

[00167] In a specific embodiment, a host cell (e.g., recombinant host cell) capable of producing a bioconjugate of an *E. coli* glucosylated O4 antigen polysaccharide covalently linked to a carrier protein further comprises a nucleotide sequence encoding PglB oligosaccharyl transferase from *Campylobacter jejuni* having the amino acid sequence of SEQ ID NO: 6, or preferably SEQ ID NO: 6 comprising the mutation N311V, or more preferably SEQ ID NO: 6 comprising the mutations Y77H and N311V.

[00168] In other specific embodiments, a host cell (e.g., recombinant host cell) capable of producing a bioconjugate of an *E. coli* O1A, O6A, or O15 antigen polysaccharide covalently linked to a carrier protein further comprises a nucleotide sequence encoding PglB oligosaccharyl transferase from *Campylobacter jejuni* having the amino acid sequence of SEQ ID NO: 6, or preferably SEQ ID NO: 6 comprising the mutations N311V, K482R, D483H, and A669V.

[00169] In a specific embodiment, a host cell (e.g., recombinant host cell) capable of producing a bioconjugate of an *E. coli* O16 antigen polysaccharide covalently linked to a carrier protein further comprises a nucleotide sequence encoding PglB oligosaccharyl transferase from *Campylobacter jejuni* having the amino acid sequence of SEQ ID NO: 6, or preferably SEQ ID NO: 6 comprising the mutations Y77H, S80R, Q287P, K289R, and N311V.

[00170] In a specific embodiment, a host cell (e.g., recombinant host cell) capable of producing a bioconjugate of an *E. coli* O75 antigen polysaccharide covalently linked to a carrier protein further comprises a nucleotide sequence encoding PglB oligosaccharyl transferase from *Campylobacter jejuni* having the amino acid sequence of SEQ ID NO: 6, or preferably SEQ ID NO: 6 comprising the mutation N311V.

[00171] In a specific embodiment, a host cell (e.g., recombinant host cell) capable of producing a bioconjugate of an *E. coli* O8, O18A, O25B, or O2 antigen polysaccharide covalently linked to a carrier protein further comprises a nucleotide sequence encoding PglB oligosaccharyl transferase from *Campylobacter jejuni* having the amino acid sequence of SEQ ID NO: 6, preferably wherein SEQ ID NO: 6 comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483, and 669.

[00172] In some embodiments, any of the host cells provided herein comprise a nucleic acid encoding a carrier protein, e.g., a protein to which the O-antigen polysaccharide(s) produced by the host cell glycosylation machinery can be attached to form a bioconjugate. The host cell can comprise a nucleic acid encoding any carrier protein known to those skilled in the art in view of the present disclosure including, but not limited to, detoxified Exotoxin A of *P. aeruginosa* (EPA), *E. coli* flagellin (FliC), CRM197, maltose binding protein (MBP), Diphtheria toxoid, Tetanus toxoid, detoxified hemolysin A of *S. aureus*, clumping factor A, clumping factor B, *E. coli* heat labile enterotoxin, detoxified variants of *E. coli* heat labile enterotoxin, Cholera toxin B subunit (CTB), cholera toxin, detoxified variants of cholera toxin, *E. coli* Sat protein, the passenger domain of *E. coli* Sat protein, *Streptococcus pneumoniae* Pneumolysin, Keyhole limpet hemocyanin (KLH), *P. aeruginosa* PcrV, outer membrane protein of *Neisseria meningitidis* (OMPC), and protein D from non-typeable *Haemophilus influenzae*.

[00173] In preferred embodiments, a host cell further comprises a nucleic acid encoding detoxified Exotoxin A of *P. aeruginosa* (EPA). Preferably, the EPA carrier protein comprises 1-10 glycosylation sites, preferably 2 to 4 glycosylation sites, most preferably 4 glycosylation sites, such as 1-10, preferably 2-4, and more preferably 4 glycosylation sites each comprising a glycosylation consensus sequence having the amino acid sequence of SEQ ID NO: 1, and more preferably having the amino acid sequence of SEQ ID NO: 2. In a specific embodiment, a host cell further comprises a nucleic acid encoding EPA-4 carrier protein comprising SEQ ID NO: 3.

[00174] In certain embodiments, the carrier proteins used in the generation of the bioconjugates by the host cells described herein comprise a “tag,” i.e., a sequence of amino acids that allows for the isolation and/or identification of the carrier protein. For example, adding a tag to a carrier protein can be useful in the purification of that protein and, hence, the purification of conjugate vaccines comprising the tagged carrier protein. Exemplary tags that can be used herein include, without limitation, histidine (HIS) tags (e.g., hexa-histidine-tag, or 6XHis-Tag), FLAG-TAG, and HA tags. In certain embodiments, the tags used herein are removable, e.g., removal by chemical agents or by enzymatic means, once they are no longer needed, e.g., after the protein has been purified. In other embodiments, the carrier protein does not comprise a tag.

[00175] In certain embodiments, the carrier proteins described herein comprise a signal sequence that targets the carrier protein to the periplasmic space of the host cell that expresses

the carrier protein. In a specific embodiment, the signal sequence is from *E. coli* DsbA, *E. coli* outer membrane porin A (OmpA), *E. coli* maltose binding protein (MalE), *Erwinia carotovorans* pectate lyase (PelB), FlgI, NikA, or *Bacillus* sp. endoxylanase (XynA), heat labile *E. coli* enterotoxin LTIIb, *Bacillus* endoxylanase XynA, or *E. coli* flagellin (FlgI). In one embodiment, the signal sequence comprises SEQ ID NO: 10. A signal sequence may be cleaved off after translocation of the protein to the periplasm and may thus no longer be present in the final carrier protein of a bioconjugate.

[00176] In certain embodiments, additional modifications can be introduced (e.g., using recombinant techniques) into the host cells described herein. For example, host cell nucleic acids (e.g., genes) that encode proteins that form part of a possibly competing or interfering glycosylation pathway (e.g., compete or interfere with one or more heterologous genes involved in glycosylation that are recombinantly introduced into the host cell) can be deleted or modified in the host cell background (genome) in a manner that makes them inactive/dysfunctional (i.e., the host cell nucleic acids that are deleted/modified do not encode a functional protein). In certain embodiments, when nucleic acids are deleted from the genome of the host cells provided herein, they are replaced by a desirable sequence, e.g., a sequence that is useful for production of an O antigen polysaccharide or bioconjugate thereof.

[00177] Exemplary genes or gene clusters that can be deleted in host cells (and, in some cases, replaced with other desired nucleic acid sequences) include genes or gene clusters of host cells involved in glycolipid biosynthesis, such as *waaL* (see, e.g., Feldman et al., 2005, *PNAS USA* 102:3016-3021), the lipid A core biosynthesis cluster (*waa*), galactose cluster (*gal*), arabinose cluster (*ara*), colonic acid cluster (*wc*), capsular polysaccharide cluster, undecaprenol-p biosynthesis genes (e.g. *uppS*, *uppP*), und-P recycling genes, metabolic enzymes involved in nucleotide activated sugar biosynthesis, enterobacterial common antigen cluster (*eca*), and prophage O antigen modification clusters like the *gtrABS* cluster or regions thereof. In a specific embodiment, the host cells described herein are modified such that they do not produce any O antigen polysaccharide other than a desired O antigen polysaccharide, e.g., glucosylated O4 antigen polysaccharide.

[00178] In a specific embodiment, the *waaL* gene is deleted or functionally inactivated from the genome of a host cell (e.g., recombinant host cell) provided herein. The terms “*waaL*” and

“*waaL* gene” refer to the O-antigen ligase gene encoding a membrane bound enzyme with an active site located in the periplasm. The encoded enzyme transfers undecaprenylphosphate (UPP)-bound O antigen to the lipid A core, forming lipopolysaccharide. Deletion or disruption of the endogenous *waaL* gene (e.g., $\Delta waaL$ strains) disrupts transfer of the O-antigen to lipid A, and can instead enhance transfer of the O-antigen to another biomolecule, such as a carrier protein.

[00179] In another specific embodiment, one or more of the *waaL* gene, *gtrA* gene, *gtrB* gene, *gtrS* gene, and the *rfb* gene cluster is deleted or functionally inactivated from the original genome of a prokaryotic host cell provided herein.

[00180] In one embodiment, a host cell used herein is *E. coli* that produces a bioconjugate of glucosylated O4 antigen polysaccharide, wherein the *waaL* gene is deleted or functionally inactivated from the genome of the host cell, and a *gtrS* gene specific to *E. coli* O4 antigen polysaccharide is inserted. In certain embodiments for production strains for bioconjugates of the glucosylated O4 O-antigen, a *gtrS* gene encoding a glucosyl transferase having at least 80% sequence identity to SEQ ID NO:4 is inserted in the place of a *gtrS* gene of the parent strain, so as to replace the *gtrS* gene in that parent strain with the one that is responsible for glucosylation of the O4 antigen. An example of such a parent strain is *E. coli* K-12 strain W3110. The *gtrA* and *gtrB* genes can be homologous to the parent strain, or alternatively one or both of these genes can be heterologous to the parent strain. Typically, and unlike the *gtrS* gene, these *gtrA* and *gtrB* genes are not specific for the O-antigen structure.

[00181] Also provided herein are methods of making recombinant host cells. Recombinant host cells produced by the methods described herein can be used to produce bioconjugates of *E. coli* O antigens. The methods comprise introducing one or more recombinant nucleic acid molecules into a cell to produce the recombinant host cell. Typically, the recombinant nucleic acid molecules are heterologous. Any method known in the art in view of the present disclosure can be used to introduce recombinant nucleic acid molecules into a host cell. Recombinant nucleic acids can be introduced into the host cells described herein using any methods known to those of ordinary skill in the art, e.g., electroporation, chemical transformation, by heat shock, natural transformation, phage transduction, and conjugation. In specific embodiments, recombinant nucleic acids are introduced into the host cells described herein using a plasmid.

For example, the heterologous nucleic acids can be expressed in the host cells by a plasmid (e.g., an expression vector). In another specific embodiment, heterologous nucleic acids are introduced into the host cells described herein using the method of insertion into the genome as for instance described in International Patent Application Publication WO 2014/037585, WO 2014/057109, or WO 2015/052344.

[00182] In one embodiment, a method of making a recombinant host cell for producing a bioconjugate of an *E. coli* glucosylated O4 antigen polysaccharide covalently linked to a carrier protein comprises introducing one or more recombinant nucleic acid molecules into a cell, preferably an *E. coli* cell, to produce the recombinant host cell. In such embodiments, the recombinant nucleic acid molecules introduced into the cell include (i) a nucleotide sequence of an *rfb* gene cluster for the *E. coli* O4 antigen polysaccharide; (ii) a nucleotide sequence encoding a glucosyl transferase having at least 80% sequence identity to SEQ ID NO: 4, wherein the glucosyl transferase is capable of modifying the *E. coli* O4 antigen polysaccharide to produce the *E. coli* glucosylated O4 antigen polysaccharide; (iii) a nucleotide sequence encoding a carrier protein; and (iv) a nucleotide sequence encoding an oligosaccharyl transferase capable of covalently linking the *E. coli* glucosylated O4 antigen polysaccharide to the carrier protein to produce the bioconjugate. In preferred embodiments, the nucleotide sequence encoding a glucosyl transferase having at least 80% sequence identity to SEQ ID NO: 4 replaces the endogenous *gtrS* gene. Deleting the endogenous *gtrS* has the advantage that it will not interfere with generation of the glucosylated O4 antigen polysaccharide structure. In certain embodiments, the nucleotide sequence of the *rfb* gene cluster for the *E. coli* O4 antigen polysaccharide replaces the endogenous *rfb* gene cluster of the parent strain that is used to make the recombinant host cell. If the cell does not yet encode *gtrA* and/or *gtrB* genes, nucleotide sequences encoding a translocase (*gtrA*) and a glycosyltransferase (*gtrB*), having at least 80% identity to SEQ ID NOs: 7 and 8, respectively, can be introduced into the cell. If the cell already encodes *gtrA* and *gtrB* genes (such as for instance the case in *E. coli* K-12 strain W3110), there is no need to introduce or change these genes.

[00183] In a specific embodiment, the glucosyl transferase (*gtrS* specific for adding glucose branch to O4 antigen) has SEQ ID NO: 4.

[00184] In a specific embodiment, the oligosaccharyl transferase is PglB from *C. jejuni*. In one such embodiment, the oligosaccharyl transferase comprises the amino acid sequence of SEQ ID NO: 6. In another such embodiment, the oligosaccharyl transferase comprises the amino acid sequence of SEQ ID NO: 6 comprising the mutation N311V. In another such embodiment, the oligosaccharyl transferase comprises the amino acid sequence of SEQ ID NO: 6 comprising the mutations Y77H and N311V.

[00185] In another specific embodiment, the carrier protein comprises at least one glycosylation site comprising a glycosylation consensus sequence having SEQ ID NO: 1, preferably SEQ ID NO: 2. In another specific embodiment, the carrier protein is EPA, preferably EPA-4, such as EPA-4 comprising SEQ ID NO: 3.

[00186] *E. coli* strains that are used routinely in molecular biology as both a tool and a model organism can for instance be used as parents for host cells in certain embodiments according to the invention. Non-limiting examples include *E. coli* K12 strains (for example, such as W1485, W2637, W3110, MG1655, DH1, DH5 α , DH10, etc.), B strains (e.g. BL-21, REL606, etc.), C strains, or W strains. In one particular embodiment, the host strain is derived from parent strain W3110. This strain can for instance be obtained from the *E. coli* Genetic Stock Center at Yale. For more information on *E. coli*, see e.g. Ecoliwiki.net.

[00187] Methods of Producing Conjugates and Bioconjugates

[00188] Also provided are methods of producing glycoconjugates of the *E. coli* O antigen polysaccharides described herein. Glycoconjugates, including bioconjugates, can be prepared *in vitro* or *in vivo*, e.g., using the recombinant host cells described herein for production.

[00189] In some embodiments, glycoconjugates can be prepared by chemical synthesis, i.e., prepared outside of host cells (*in vitro*). For example, an *E. coli* O antigen polysaccharide can be conjugated to carrier proteins using methods known to those of ordinary skill in the art, including by means of using activation reactive groups in the polysaccharide/oligosaccharide as well as the carrier protein. See, e.g., Pawlowski et al., 2000, *Vaccine* 18:1873-1885; and Robbins, et al., 2009, *Proc Natl Acad Sci USA* 106:7974-7978), the disclosures of which are herein incorporated by reference. Such approaches comprise extraction of antigenic polysaccharides/oligosaccharides from host cells, purifying the polysaccharides/oligosaccharides, chemically

activating the polysaccharides/oligosaccharides, and conjugating the polysaccharides/oligosaccharides to a carrier protein.

[00190] In some embodiments, the host cells described herein can be used to produce bioconjugates comprising an *E. coli* O antigen polysaccharide covalently linked to a carrier protein. Methods of producing such bioconjugates using host cells are known in the art. See, e.g., WO 2003/074687 and WO 2006/119987. Such methods comprise culturing any of the recombinant host cells described herein under conditions for production of the bioconjugate. Bioconjugates can be isolated, separated, and/or purified from recombinant host cells using any method known in the art in view of the present disclosure. For example, bioconjugates can be purified by any method known in the art for purification of a protein, for instance, by chromatography (e.g., ion exchange, anionic exchange, affinity, and sizing column chromatography), centrifugation, differential solubility, or by any other standard technique for the purification of proteins. See, e.g., methods described in WO 2009/104074. Further, the bioconjugates can be fused to heterologous polypeptide sequences to facilitate purification. The actual conditions used to purify a particular bioconjugate will depend, in part, on factors such as net charge, hydrophobicity, and/or hydrophilicity of the bioconjugate, and will be apparent to those skilled in the art. Preparation of bioconjugates for O1A, O2, O6A, and O25B, as well as vaccine compositions comprising these, have for instance been described in WO 2015/124769 and in WO 2017/035181.

[00191] Also provided are bioconjugates produced by the methods described herein, i.e., using the recombinant host cells described herein.

[00192] In some embodiments, a method of preparing a bioconjugate of an *E. coli* O-antigen polysaccharide covalently linked to a carrier protein comprises: (i) providing a recombinant host cell comprising (a) nucleotide sequence of an *rfb* gene cluster for the O-antigen polysaccharide; (b) a nucleotide sequence encoding a carrier protein, preferably EPA, comprising at least one glycosylation site comprising a glycosylation consensus sequence having SEQ ID NO: 1, preferably SEQ ID NO: 2, and more preferably comprising four glycosylation sites each comprising a glycosylation consensus sequence having SEQ ID NO: 2; and (c) nucleotide sequence encoding an oligosaccharyl transferase, for instance PglB oligosaccharyl transferase or variant thereof.

[00193] In certain embodiments, *E. coli* O-antigen polysaccharides produced using the recombinant host cells described herein are covalently bound to the carrier protein at a particular polysaccharide to protein ratio by weight (w/w). This ratio of amount of O-antigen polysaccharide by weight covalently bound to the carrier protein by weight is referred to as the “glycan/protein ratio” or “polysaccharide/protein ratio” or “PS/protein ratio”. In some embodiments, the O-antigen polysaccharide is covalently bound to the carrier protein at a polysaccharide to protein (w/w) ratio of about 1: 20 to 20:1, preferably 1:10 to 10:1, more preferably 1:3 to 3:1. In certain non-limiting embodiments for bioconjugates described herein, glycan/protein ratio is about 0.1 to 0.5, such as 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, or 0.5. In such embodiments, the weight ratio of the O-antigen polysaccharide: protein is about 1:10 to 1:2, such as 1:10: 1:9: 1:8, 1:7, 1:6, 1:5, 1:4, 1:3, or 1:2, depending on the particular O-antigen serotype. In certain embodiments the glycan/protein ratio is from about 0.15 to about 0.45. In general, a higher glycan/protein ratio of O-antigen polysaccharide to carrier protein is preferred, because a high amount of carrier protein can lead to immunological interference in some instances. Also, a higher glycan/protein ratio would help getting sufficient O-antigen polysaccharide dosed in the form of bioconjugate, while keeping the amount of carrier protein relatively low, which is especially beneficial for multivalent compositions where multiple serotypes are to be covered by the composition, e.g. compositions comprising bioconjugates from at least 4 different O-antigens, at least 5 different O-antigens, at least 6 different O-antigens, at least 7 different O-antigens, at least 8 different O-antigens, at least 9 different O-antigens, at least 10 different O-antigens, etc.

[00194] A glycan/protein ratio of a conjugate according to the invention can be determined by determining the protein amount and the glycan amount. Protein amount can be determined by measurement of UV absorbance at 280 nm (A₂₈₀). Glycan amount can be determined based on ion chromatography with pulsed amperometric detection (IC-PAD) of a sugar in the repeat unit (e.g. of Man for O8 in Table 1, and of GlcNAc for the other glycans in Table 1), after which the structural information of the repeat unit can be used to calculate the total glycan amount (e.g. the repeat unit of O1A has a molar mass of 845 Da and one mole of such a repeat unit contains one mole of GlcNAc, enabling calculation of the total glycan amount when the amount of GlcNAc has been determined by IC-PAD).

[00195] In some embodiments, a bioconjugate of an *E. coli* O25B antigen polysaccharide covalently linked to a carrier protein produced using a recombinant host cell according to the cells and methods described herein has a certain degree of acetylation at position 2 of the L-Rh sugar. The degree of O-acetylation of O25B antigen polysaccharide in a bioconjugate is preferably at least 30%, preferably at least 50%, such as at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100%.

[00196] Similarly, the degree of O-acetylation of an *E. coli* O16 antigen polysaccharide in a bioconjugate is preferably at least 30%, preferably at least 50%, such as at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100%.

[00197] In specific embodiments, a method of preparing a bioconjugate of an O-antigen polysaccharide comprises providing a recombinant host cell comprising nucleic acid sequence encoding a particular oligosaccharyl transferase enzyme, particularly a PglB oligosaccharyl transferase or variant thereof, depending on the O-antigen polysaccharide bioconjugate to be produced. The particular oligosaccharyl transferase enzyme variant may impact the yield of bioconjugate produced by the host cell. Typically, a higher yield is preferred, since the yield will impact the costs for producing a specific bioconjugate, which is especially important for multivalent compositions comprising several different bioconjugates. In some embodiments, the method further comprises isolating the bioconjugate from the recombinant host cell.

[00198] In one particular embodiment, when the O- antigen is O1A, O6A, or O15 antigen polysaccharide, the PglB oligosaccharyl transferase comprises the amino acid mutations of N311V, K482R, D483H, and A669V, wherein the amino acid mutations are relative to the wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[00199] In another particular embodiment, when the O-antigen is glucosylated O4 antigen polysaccharide, the PglB oligosaccharyl transferase comprises the amino acid mutation N311V, or the amino acid mutations of Y77H and N311V, wherein the amino acid mutations are relative to the wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[00200] In another particular, embodiment, when the O-antigen is O16 antigen polysaccharide, the PglB oligosaccharyl transferase comprises the amino acid mutations of Y77H, S80R, Q287P, K289R, and N311V, wherein the amino acid mutations are relative to the wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[00201] In another particular embodiment, when the O-antigen is O75 antigen polysaccharide, the PglB oligosaccharyl transferase comprises the amino acid mutation of N311V, wherein the amino acid mutations are relative to the wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[00202] In another particular embodiment, when the O-antigen is O8, O18A, O25B, or O2 antigen polysaccharide, the PglB oligosaccharyl transferase comprises the amino acid sequence of SEQ ID NO: 6, wherein SEQ ID NO: 6 comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483, and 669. In certain embodiments thereof, the PglB oligosaccharyl transferase comprises the amino acid sequence of SEQ ID NO: 6.

[00203] In some embodiments, the carrier protein is selected from the group consisting of detoxified Exotoxin A of *P. aeruginosa* (EPA), *E. coli* flagellin (FliC), CRM197, maltose binding protein (MBP), Diphtheria toxoid, Tetanus toxoid, detoxified hemolysin A of *S. aureus*, clumping factor A, clumping factor B, *E. coli* heat labile enterotoxin, detoxified variants of *E. coli* heat labile enterotoxin, Cholera toxin B subunit (CTB), cholera toxin, detoxified variants of cholera toxin, *E. coli* Sat protein, the passenger domain of *E. coli* Sat protein, Streptococcus pneumoniae Pneumolysin, Keyhole limpet hemocyanin (KLH), *P. aeruginosa* PcrV, outer membrane protein of *Neisseria meningitidis* (OMPC), and protein D from non-typeable *Haemophilus influenzae*.

[00204] In certain embodiments, the carrier protein is detoxified exotoxin A of *Pseudomonas aeruginosa* (EPA). Preferably, the EPA carrier protein comprises 1-10, preferably 2-4, more preferably 4 glycosylation sites. Preferably, each glycosylation site comprises a glycosylation consensus sequence having the amino acid sequence of SEQ ID NO: 2. In a specific embodiment, a host cell comprises a nucleic acid encoding EPA-4 carrier protein comprising SEQ ID NO: 3.

[00205] In certain embodiments, the recombinant host cell is an *E. coli* cell, e.g., an *E. coli* K-12 strain, such as strain W3110.

[00206] Also provided herein are bioconjugates of O-antigen polysaccharides produced using recombinant host cells encoding the oligosaccharyl transferase enzymes per the O-antigen/PglB oligosaccharyl transferase pairings indicated above. Also provided are compositions comprising

such bioconjugates. In certain embodiments, a composition comprises at least 2, preferably at least 3, more preferably at least 5, still more preferably at least 7 of such bioconjugates.

[00207] In some embodiments, bioconjugates of O-antigen polysaccharides produced by recombinant host cells encoding the oligosaccharyl transferase enzymes per the O-antigen/PglB oligosaccharyl transferase pairings indicated above preferably have one or more of the preferred attributes described herein, e.g., glycan/protein ratio and/or amount or ratio of multi-glycosylated carrier protein.

EMBODIMENTS

[00208] Embodiment 1 is a method of preparing a bioconjugate of an *E. coli* O_x antigen polysaccharide covalently linked to a carrier protein, the method comprising:

(i) providing a recombinant host cell comprising:

- a. a nucleotide sequence of an *rfb* gene cluster for the O_x-antigen polysaccharide;
- b. a nucleotide sequence encoding the carrier protein comprising at least one glycosylation site comprising a glycosylation consensus sequence having SEQ ID NO: 1, preferably having SEQ ID NO: 2; and
- c. a nucleotide sequence encoding an oligosaccharyl transferase PglB_y; and

(ii) culturing the recombinant host cell under conditions for production of the bioconjugate; wherein:

when the O_x-antigen is O1A antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V;

when the O_x-antigen is glucosylated O4 antigen polysaccharide, the PglB_y comprises the amino acid mutation N311V or the amino acid mutations Y77H and N311V, and the recombinant host cell further comprises a sequence encoding a glucosyltransferase GtrS having at least 80% identity to SEQ ID NO: 4 and being capable of modifying an *E. coli* O4 antigen polysaccharide by addition of glucose to produce the *E. coli* glucosylated O4 antigen polysaccharide, and nucleotide sequences encoding a translocase GtrA and a glycosyltransferase GtrB having at least 80% sequence identity to SEQ ID NOs: 7 and 8 respectively, wherein the translocase is capable of translocating bactoprenol linked glucose and the glycosyltransferase is capable of glucosylating bactoprenol;

when the O_x-antigen is O6A antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V;

when the O_x-antigen is O8 antigen polysaccharide, the PglB_y comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483 and 669;

when the O_x-antigen is O15 antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V;

when the O_x-antigen is O16 antigen polysaccharide, the PglB_y comprises the amino acid mutations of Y77H, S80R, Q287P, K289R, and N311V;

when the O_x-antigen is O18A antigen polysaccharide, the PglB_y comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483 and 669; and

when the O_x-antigen is O75 antigen polysaccharide, the PglB_y comprises the amino acid mutation of N311V;

wherein in each case the amino acid mutations are relative to the wild-type PglB having the amino acid sequence of SEQ ID NO: 6, and

wherein the O1A, glucosylated O4, O6A, O8, O15, O16, O18A, and O75 antigen polysaccharides have the structures of Formulas (O1A), (O4-Glc+), (O6A), (O8), (O15), (O16), (O18A), and (O75), respectively, as shown in Table 1, and each n is independently an integer of 1 to 100, preferably 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20.

[00209] Embodiment 2 is the method of embodiment 1, wherein the O_x-antigen is O1A antigen polysaccharide, and the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[00210] Embodiment 3 is the method of embodiment 1, wherein the O_x-antigen is glucosylated O4 antigen polysaccharide, and the PglB_y comprises the amino acid mutation N311V or the amino acid mutations Y77H and N311V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[00211] Embodiment 4 is the method of embodiment 3, wherein the recombinant host cell further comprises a sequence encoding a GtrS having the amino acid sequence of SEQ ID NO: 4, and nucleotide sequences encoding a GtrA and a GtrB having the amino acid sequences of SEQ ID NOs: 7 and 8, respectively.

[00212] Embodiment 5 is the method of embodiment 1, wherein the O_x-antigen is O6A antigen polysaccharide, and the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[00213] Embodiment 6 is the method of embodiment 1, wherein the O_x-antigen is O8 antigen polysaccharide, and the PglB_y comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483 and 669 relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[00214] Embodiment 7 is the method of embodiment 1, wherein the O_x-antigen is O15 antigen polysaccharide, and the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[00215] Embodiment 8 is the method of embodiment 1, wherein the O_x-antigen is O16 antigen polysaccharide, and the PglB_y comprises the amino acid mutations of Y77H, S80R, Q287P, K289R, and N311V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[00216] Embodiment 9 is the method of embodiment 1, wherein the O_x-antigen is O18A antigen polysaccharide, and the PglB_y comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483 and 669 relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[00217] Embodiment 10 is the method of embodiment 1, wherein the O_x-antigen is O75 antigen polysaccharide, and the PglB_y comprises the amino acid mutation of N311V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[00218] Embodiment 11 is a method of preparing a bioconjugate of an *E. coli* O_x antigen polysaccharide covalently linked to a carrier protein, the method comprising:

(i) providing a recombinant host cell comprising:

(a) a nucleotide sequence of an *rfb* gene cluster for the O_x-antigen polysaccharide;

(b) a nucleotide sequence encoding the carrier protein comprising at least one glycosylation site comprising a glycosylation consensus sequence having SEQ ID NO: 1, preferably having SEQ ID NO: 2; and

(c) a nucleotide sequence encoding an oligosaccharyl transferase PglB_y; and

(ii) culturing the recombinant host cell under conditions for production of the bioconjugate,

wherein the PglB_y comprises the amino acid mutation N311V relative to the wild-type PglB having the amino acid sequence of SEQ ID NO: 6,

wherein the O_x- antigen is O1A antigen polysaccharide, glucosylated O4 antigen polysaccharide, O6A antigen polysaccharide, O15 antigen polysaccharide, O16 antigen polysaccharide, or O75 antigen polysaccharide, and when the O_x-antigen is glucosylated O4 antigen polysaccharide, the recombinant host cell further comprises a sequence encoding a glucosyltransferase GtrS having at least 80% identity to SEQ ID NO: 4 and being capable of modifying an *E. coli* O4 antigen polysaccharide by addition of glucose to produce the *E. coli* glucosylated O4 antigen polysaccharide, and nucleotide sequences encoding a translocase GtrA and a glycosyltransferase GtrB having at least 80% sequence identity to SEQ ID NOs: 7 and 8, respectively, wherein the translocase is capable of translocating bactoprenol linked glucose and the glycosyltransferase is capable of glucosylating bactoprenol, and

wherein the O1A, glucosylated O4, O6A, O15, O16, and O75 antigen polysaccharides have the structures of Formulas (O1A), (O4-Glc+), (O6A), (O15), (O16), and (O75), respectively, as shown in Table 1, and each n is independently an integer of 1 to 100, preferably 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20.

[00219] Embodiment 12 is the method of any one of embodiments 1 to 11, further comprising isolating the bioconjugate from the recombinant host cell.

[00220] Embodiment 13 is the method of any one of embodiments 1 to 12, wherein the carrier protein is selected from the group consisting of detoxified Exotoxin A of *P. aeruginosa* (EPA), *E. coli* flagellin (FliC), CRM197, maltose binding protein (MBP), Diphtheria toxoid, Tetanus toxoid, detoxified hemolysin A of *S. aureus*, clumping factor A, clumping factor B, *E. coli* heat labile enterotoxin, detoxified variants of *E. coli* heat labile enterotoxin, Cholera toxin B subunit (CTB), cholera toxin, detoxified variants of cholera toxin, *E. coli* Sat protein, the passenger domain of *E. coli* Sat protein, *Streptococcus pneumoniae* Pneumolysin, Keyhole limpet hemocyanin (KLH), *P. aeruginosa* PcrV, outer membrane protein of *Neisseria meningitidis* (OMPc), and protein D from non-typeable *Haemophilus influenzae*.

[00221] Embodiment 14 is the method of embodiment 13, wherein the carrier protein is detoxified exotoxin A of *Pseudomonas aeruginosa* (EPA).

[00222] Embodiment 15 is the method of embodiment 14, wherein the EPA carrier protein comprises 1-10, preferably 2-4, more preferably 4, of the glycosylation sites.

[00223] Embodiment 16 is the method of embodiment 15, wherein each glycosylation site comprises a glycosylation consensus sequence having SEQ ID NO: 2.

[00224] Embodiment 17 is the method of embodiment 16, wherein the EPA carrier protein comprises SEQ ID NO: 3.

[00225] Embodiment 18 is the method of any one of embodiments 1-17, wherein the recombinant host cell is an *E. coli* cell, e.g. an *E. coli* K-12 strain, such as strain W3110.

[00226] Embodiment 19 is a bioconjugate produced by the method of any one of embodiments 1-18.

[00227] Embodiment 20 is a composition comprising a bioconjugate of embodiment 19.

[00228] Embodiment 21 is a composition comprising at least 2, preferably at least 3, more preferably at least 5, still more preferably at least 7 bioconjugates of embodiment 19.

[00229] Embodiment 22 is a composition of embodiment 20 or 21, comprising a bioconjugate of *E. coli* glucosylated O4 antigen polysaccharide covalently linked to a carrier protein, wherein the glucosylated O4 antigen polysaccharide has the structure of Formula (O4-Glc+) as shown in Table 1, and n is an integer of 1 to 100, preferably 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20.

[00230] Embodiment 23 is a composition of any one of embodiments 20 to 22, further comprising at least a bioconjugate of *E. coli* O25B antigen polysaccharide covalently linked to a carrier protein, wherein the O25B antigen polysaccharide has the structure of Formula (O25B) as shown in Table 1, and n is an integer of 1 to 100, preferably 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20.

[00231] Embodiment 24 is a composition of any one of embodiments 20 to 23, further comprising at least a bioconjugate of *E. coli* O2 antigen polysaccharide covalently linked to a carrier protein, wherein the O2 antigen polysaccharide has the structure of Formula (O2) as shown in Table 1, and n is an integer of 1 to 100, preferably 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20.

[00232] Embodiment 25 is a composition of any one of embodiments 20 to 24, comprising:

- (i) bioconjugate of *E. coli* O1A antigen polysaccharide covalently coupled to a carrier protein,
- (ii) bioconjugate of *E. coli* O2 antigen polysaccharide covalently coupled to a carrier protein, (iii)

bioconjugate of *E. coli* glucosylated O4 antigen polysaccharide covalently coupled to a carrier protein, (iv) bioconjugate of *E. coli* O6A antigen polysaccharide covalently coupled to a carrier protein, (v) bioconjugate of *E. coli* O8 antigen polysaccharide covalently coupled to a carrier protein, (vi) bioconjugate of *E. coli* O15 antigen polysaccharide covalently coupled to a carrier protein, (vii) bioconjugate of *E. coli* O16 antigen polysaccharide covalently coupled to a carrier protein, (viii) bioconjugate of *E. coli* O25B antigen polysaccharide covalently coupled to a carrier protein, and (ix) bioconjugate of *E. coli* O75 antigen polysaccharide covalently coupled to a carrier protein, wherein the O1A, O2, glucosylated O4, O6A, O8, O15, O16, O25B, and O75 antigen polysaccharides have the structures of Formulas (O1A), (O2), (O4-Glc+), (O6A), (O8), (O15), (O16), (O25B), and (O75), respectively, as shown in Table 1, and each n is independently an integer of 1 to 100, preferably 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20.

[00233] Embodiment 26 is a composition of embodiment 25, further comprising: (x) bioconjugate of *E. coli* O18A antigen polysaccharide covalently coupled to a carrier protein, wherein the O18A antigen polysaccharide has the structure of Formula (O18A) as shown in Table 1, and n is an integer of 1 to 100, preferably 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20.

[00234] Embodiment 27 is a composition of any one of embodiments 20 to 26, wherein the composition is an immunogenic composition.

[00235] Embodiment 28 is a method of vaccinating a subject against *E. coli*, in particular extra-intestinal pathogenic *E. coli* (ExPEC), comprising administering to the subject the bioconjugate of embodiment 19, or the composition or immunogenic composition of any one of embodiments 20 to 27.

[00236] Embodiment 29 is the bioconjugate of embodiment 19, or the composition or immunogenic composition of any one of embodiments 20 to 27 for use in vaccination against extra-intestinal pathogenic *E. coli* (ExPEC).

[00237] Embodiment 30 is a recombinant host cell for preparing a bioconjugate of an *E. coli* O_x antigen polysaccharide covalently linked to a carrier protein, the recombinant host cell comprising:

- (a) a nucleotide sequence of an *rfb* gene cluster for the O_x-antigen polysaccharide;
- (b) a nucleotide sequence encoding the carrier protein comprising at least one glycosylation site comprising a glycosylation consensus sequence having SEQ ID NO: 1, preferably having SEQ

ID NO: 2; and

(c) a nucleotide sequence encoding an oligosaccharyl transferase PglB_y;

wherein:

when the O_x-antigen is O1A antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V;

when the O_x-antigen is glucosylated O4 antigen polysaccharide, the PglB_y comprises the amino acid mutation N311V or the amino acid mutations Y77H and N311V, and the recombinant host cell further comprises a sequence encoding a glucosyltransferase GtrS having at least 80% identity to SEQ ID NO: 4 and being capable of modifying an *E. coli* O4 antigen polysaccharide by addition of glucose to produce the *E. coli* glucosylated O4 antigen polysaccharide, and nucleotide sequences encoding a translocase GtrA and a glycosyltransferase GtrB having at least 80% sequence identity to SEQ ID NOs: 7 and 8 respectively, wherein the translocase is capable of translocating bactoprenol linked glucose and the glycosyltransferase is capable of glucosylating bactoprenol;

when the O_x-antigen is O6A antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V;

when the O_x-antigen is O8 antigen polysaccharide, the PglB_y comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483 and 669;

when the O_x-antigen is O15 antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V;

when the O_x-antigen is O16 antigen polysaccharide, the PglB_y comprises the amino acid mutations of Y77H, S80R, Q287P, K289R, and N311V;

when the O_x-antigen is O18A antigen polysaccharide, the PglB_y comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483 and 669; and

when the O_x-antigen is O75 antigen polysaccharide, the PglB_y comprises the amino acid mutation of N311V;

wherein in each case the amino acid mutations are relative to the wild-type PglB having the amino acid sequence of SEQ ID NO: 6, and

wherein the O1A, glucosylated O4, O6A, O8, O15, O16, O18A, and O75 antigen polysaccharides have the structures of Formulas (O1A), (O4-Glc+), (O6A), (O8), (O15), (O16),

(O18A), and (O75), respectively, as shown in Table 1, and each n is independently an integer of 1 to 100, preferably 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20.

[00238] Embodiment 31 is the recombinant host cell of embodiment 30, wherein the O_x-antigen is O1A antigen polysaccharide, and the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[00239] Embodiment 32 is the recombinant host cell of embodiment 30, wherein the O_x-antigen is glucosylated O4 antigen polysaccharide, and the PglB_y comprises the amino acid mutation N311V or the amino acid mutations Y77H and N311V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[00240] Embodiment 33 is the recombinant host cell of embodiment 32, wherein the recombinant host cell further comprises a sequence encoding a GtrS having the amino acid sequence of SEQ ID NO: 4, and nucleotide sequences encoding a GtrA and a GtrB having the amino acid sequences of SEQ ID NOs: 7 and 8, respectively.

[00241] Embodiment 34 is the recombinant host cell of embodiment 30, wherein the O_x-antigen is O6A antigen polysaccharide, and the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[00242] Embodiment 35 is the recombinant host cell of embodiment 30, wherein the O_x-antigen is O8 antigen polysaccharide, and the PglB_y comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483 and 669 relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[00243] Embodiment 36 is the recombinant host cell of embodiment 30, wherein the O_x-antigen is O15 antigen polysaccharide, and the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[00244] Embodiment 37 is the recombinant host cell of embodiment 30, wherein the O_x-antigen is O16 antigen polysaccharide, and the PglB_y comprises the amino acid mutations of Y77H, S80R, Q287P, K289R, and N311V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[00245] Embodiment 38 is the recombinant host cell of embodiment 30, wherein the O_x-antigen is O18A antigen polysaccharide, and the PglB_y comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483 and 669 relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[00246] Embodiment 39 is the recombinant host cell of embodiment 30, wherein the O_x-antigen is O75 antigen polysaccharide, and the PglB_y comprises the amino acid mutation of N311V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

[00247] Embodiment 40 is the recombinant host cell of any one of embodiments 30 to 39, wherein the carrier protein is selected from the group consisting of detoxified Exotoxin A of *P. aeruginosa* (EPA), *E. coli* flagellin (FliC), CRM197, maltose binding protein (MBP), Diphtheria toxin, Tetanus toxin, detoxified hemolysin A of *S. aureus*, clumping factor A, clumping factor B, *E. coli* heat labile enterotoxin, detoxified variants of *E. coli* heat labile enterotoxin, Cholera toxin B subunit (CTB), cholera toxin, detoxified variants of cholera toxin, *E. coli* Sat protein, the passenger domain of *E. coli* Sat protein, *Streptococcus pneumoniae* Pneumolysin, Keyhole limpet hemocyanin (KLH), *P. aeruginosa* PcrV, outer membrane protein of *Neisseria meningitidis* (OMPC), and protein D from non-typeable *Haemophilus influenzae*.

[00248] Embodiment 41 is the recombinant host cell of any one of embodiments 30-40, wherein the carrier protein is detoxified exotoxin A of *Pseudomonas aeruginosa* (EPA).

[00249] Embodiment 42 is the recombinant host cell of embodiment 41, wherein the EPA carrier protein comprises 1-10, preferably 2-4, more preferably 4, of the glycosylation sites.

[00250] Embodiment 43 is the recombinant host cell of embodiment 42, wherein each glycosylation site comprises a glycosylation consensus sequence having SEQ ID NO: 2.

[00251] Embodiment 44 is the recombinant host cell of embodiment 43, wherein the EPA carrier protein comprises SEQ ID NO: 3.

[00252] Embodiment 45 is the recombinant host cell of any one of embodiments 30 to 44, wherein the recombinant host cell is an *E. coli* cell, e.g. an *E. coli* K-12 strain, such as strain W3110.

[00253] Embodiment 46 is a bioconjugate according to embodiment 19, wherein the bioconjugate is a bioconjugate of *E. coli* glucosylated O4 antigen polysaccharide covalently linked to a carrier protein.

[00254] Embodiment 47 is a bioconjugate according to embodiment 46, wherein the carrier protein is an EPA carrier protein comprising SEQ ID NO: 3.

[00255] Embodiment 48 is a bioconjugate according to embodiment 46 or 47, wherein the glucosylated O4 antigen polysaccharide has the structures of Formula (O4-Glc+) as shown in Table 1, and n is an integer of 5 to 40.

[00256] Embodiment 49 is a composition comprising a bioconjugate according to any one of embodiments 46-48.

[00257] Embodiment 50 is a composition according to embodiment 49, further comprising one or more conjugates each comprising an *E. coli* antigen polysaccharide covalently coupled to a carrier protein.

[00258] Embodiment 51 is a composition according to embodiment 50, wherein the one or more conjugates comprise *E. coli* antigen polysaccharide of one or more of the following *E. coli* serotypes: O1A, O2, O6A, O8, O15, O16, O18A, O25B, and O75, wherein the O1A, O2, O6A, O8, O15, O16, O25B, and O75 antigen polysaccharides have the structures of Formulas (O1A), (O2), (O6A), (O8), (O15), (O16), (O18A), (O25B), and (O75), respectively, as shown in Table 1, and each n is independently an integer of 1 to 100, preferably 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20.

[00259] Embodiment 52 is a composition according to embodiment 51, comprising conjugates of *E. coli* serotypes: O1A, O2, O6A, O8, O15, O16, O18A, O25B, and O75.

[00260] Embodiment 53 is a composition according to embodiment 52, wherein each of the conjugates is a bioconjugate.

EXAMPLES

[00261] The following examples of the invention are to further illustrate the nature of the invention. It should be understood that the following examples do not limit the invention and the scope of the invention is to be determined by the appended claims.

Example 1: Epidemiological data of *E.coli* infections

[00262] To determine the O-serotype distribution of bacteremia-causing *E. coli*, global surveillance studies were performed. Between 2011 and 2017, more than 3200 *E. coli* bloodstream isolates were collected from patients \geq 60 years of age hospitalized in countries within North America, Europe, the Asia-Pacific region, and South America. Each strain was analyzed for O antigen serotype using classical agglutination techniques and sequence-based O-genotyping. See Table 2.

[00263] Isolated human blood samples were analyzed to determine the identity of pathogens therein and their antibiotic resistance patterns. *E. coli* isolates were obtained from the samples following the analysis. *E. coli* identity was verified by MALDI-TOF MS. Further analysis on the *E. coli* isolates was performed using an antisera-based agglutination assay to determine their O-antigen serotype (DebRoy et al. (2011) Animal health research reviews / Conference of Research Workers in Animal Diseases 12, 169-185). Isolates un-typeable by the agglutination method, were further analyzed by whole-genome sequencing followed by O-genotyping based on O-serotype specific *wzy* and *wzx* gene sequences.

[00264] **Table 2:** distribution of the most common bacteremia-associated *E. coli* O-serotypes from a collection of 3217 blood isolates collected globally between 2011 and 2017, based on O-serotyping by agglutination plus O-genotyping of isolates un-typeable by agglutination. Subjects were hospitalized in the following countries: USA, Canada, Argentina, Brazil, UK, Germany, Spain, Italy, The Netherlands, France, Japan, Thailand, South Korea and Australia.

O-serotype	Prevalence n (%)
O25	737 (22.9%)
O2	268 (8.3%)
O6	261 (8.1%)
O1	255 (7.9%)
O75	145 (4.5%)
O15	110 (3.4%)
O8	104 (3.2%)
O16	103 (3.2%)
O4	96 (3.0%)
O18	91 (2.8%)

[00265] Stratification of on geographical location in the global set of bacteremia-associated *E. coli* showed a prevalence of the top 10 O-serotypes independent of location, suggesting these to be the predominant O-serotypes globally associated with bacteremia-causing *E. coli*.

[00266] In the global set of bacteremia-associated multi-drug resistant *E. coli* isolates (n=345), i.e. those strains that are resistant to at least three classes of clinically relevant antimicrobial drugs, the prevalence of the top 10 O-serotypes is 75.4%.

[00267] All information from epidemiology analysis taken together, the 10 predominant O-serotypes could cover an estimated 60-80% of *E. coli*-associated bacteremia infections, assuming coverage of subportions of the un-typeable strains.

[00268] A multivalent vaccine covering a significant proportion of bacteremia-causing *E. coli* serotypes would be very useful. The O-serotypes of Table 2 would thus be good candidates for an O-antigen based multivalent vaccine. Such a vaccine could beneficially be prepared using bioconjugation technology.

[00269] One of the serotypes in the top-10 (Table 2) is O4. It would thus be beneficial to prepare a bioconjugate vaccine that includes O-antigen polysaccharide of *E. coli* serotype O4 coupled to a carrier protein.

Example 2: Characterization of Contemporary O4 Clinical Isolates for Genes Encoding O-antigen Modifying Enzymes

[00270] Two variants of *E. coli* O4 antigen polysaccharide have been described (see, e.g. Jann B, et al., 1993, Carbohydr. Res. 248: 241-250), one having an unbranched structure (structure shown as (O4-Glc-) in Table 1) and another variant substituted with an additional glucose side-branch (structure shown as (O4-Glc+) in Table 1). The proportion in which these two variants are found in contemporary clinical isolates was not known. Although both variants react with O4 antisera, it was also not known whether immunological differences between these variants exist. Moreover, an enzyme responsible for attaching the glucose side-branch to generate the (O4-Glc+) antigen polysaccharide was hitherto not identified, and a putative coding sequence thereof is likely residing outside the O4 *rfb* gene cluster.

[00271] A set of 32 agglutination-confirmed *E. coli* O4 clinical isolates originally isolated during the period of 2011-2012 from subjects in the United States and the European Union were

subjected to whole genome sequence analysis. Extracted *rfb* gene cluster sequences from the 32 sequenced O4 isolates were aligned with those of the reference strain and compared at the nucleotide level. Except for some naturally occurring single nucleotide polymorphisms, the characterized isolates all displayed an *rfb* cluster that was identical to the O4 reference strain, indicating that *E. coli* O4 strains, independent of their Glc-branching status, carry an identical *rfb* gene cluster. Thus, to generate the *E. coli* O4-Glc+ antigen polysaccharide, a gene with unknown sequence that encodes an *E. coli* O4-specific branching enzyme and that must reside somewhere outside of the *E. coli* O4 *rfb* gene cluster is likely needed. The sequence of this unknown gene needs to be identified and employed if one wants to produce bioconjugates with the *E. coli* O4-Glc+ antigen polysaccharides in a strain that would otherwise only produce bioconjugates with *E. coli* O4-Glc- antigen polysaccharides.

[00272] The whole-genome sequence data were then analyzed for the presence of genes outside of the *rfb* gene cluster that may encode O-antigen modifying enzymes. Homologs of *gtrAB* in *Shigella flexneri* were first identified in *E. coli* O4. An open reading frame downstream of *gtrAB* in *E. coli* was then putatively identified as the *E. coli* O4-specific gene *gtrS*, that could encode the putative *E. coli* O4 specific branching enzyme GtrS responsible for adding a glucose branch to the *E. coli* O4 antigen.

[00273] The amino acid sequence of the O4 specific GtrS enzyme is provided as SEQ ID NO: 4. An exemplary nucleic acid sequence encoding this protein is provided as SEQ ID NO: 5.

[00274] Of the characterized *E. coli* O4 isolates, approximately 80% were found to carry the here identified *gtrS* gene (26 out of 32). Prevalence of the *E. coli* O4-specific *gtrS* sequence was also determined by PCR using sequence specific primers in an independent set of 20 agglutination-confirmed *E. coli* O4 clinical isolates isolated during the period of 2014-2016 from subjects in the United States and the European Union. This analysis demonstrated that 17 out of 20 isolates carried the O4 *gtrS* sequence, which corresponds to a prevalence of 85%.

Example 3: Cloning of O4 gtrS into E. coli W3110, Production and structural confirmation of Glc-Modified O4 Bioconjugates

[00275] To test whether bioconjugates comprising O4-antigen polysaccharide modified with a branching glucose could be prepared, *E. coli* O4-antigen EPA bioconjugate production strains

with the putative branching enzyme were constructed. For this, the endogenous *O16-gtrS* gene was substituted by the putative *O4-gtrS* gene (SEQ ID NO: 5, see Example 2) and the O16 *rfb* cluster was replaced with the O4 *rfb* cluster in *E. coli* strain W3110 Δ_{wzzE-wecG} Δ_{waaL} Δ_{wbbI-J-K} by homologous recombination. Alternatively, in some strains, the O4 *rfb* cluster was encoded on a plasmid.

[00276] Subsequently, plasmids encoding a detoxified exotoxin A of *Pseudomonas aeruginosa* (EPA) carrier protein (a variant either having 2 or 4 consensus glycosylation sites, referred to as ‘EPA-2’ and ‘EPA-4’, respectively), and oligosaccharyl transferase PglB were introduced into the strains. O4-EPA bioconjugates modified with Glc were produced by growing the *E. coli* production strains in bioreactor cultures, and induction of PglB and EPA expression by IPTG and arabinose, respectively. The O4-EPA bioconjugates were extracted from the biomass periplasmic extract.

[00277] To confirm the detailed polysaccharide composition and linkage of the O4-EPA bioconjugates, multiple NMR experiments were performed on the bioconjugates having EPA-4 carrier protein (data not shown). The assignments obtained agreed with literature published (Jansson, P.E., et al., 1984, Carbohydr. Res. 134(2): 283-291; Jann B, et al., 1993, Carbohydr. Res. 248: 241-250). The 1D spectrum recorded at 313K showed a large HOD signal and small sharp signals from the O4 pentasaccharide RU with five anomeric, two NAc and two H6 signals (Rha and FucNAc).

[00278] The 1D proton assignments were confirmed by use of 2D proton-proton and proton-carbon correlation NMR experiments. First, 2D TOCSY (120 ms) experiments demonstrated the expected cross peaks from H1 and H6 (for Rha and FucNAc) for the O4 pentasaccharide RU and small peaks from the terminal RU and EPA. In the methyl region, TOCSY showed cross peaks from H6 to H1 for α-Rha and H6 to H5 for α-FucNAc for the O4 RU. Other peaks observed were from EPA amino acids and terminal Rha (tRha). Second, a carbon NMR spectrum contained well-dispersed and diagnostic single peaks for the O4 RU. The carbons were profiled indirectly through the attached protons by use of the HSQC experiment. The HSQC-DEPT experiment gave inverted peaks for CH₂ groups. The HSQC gave cross peaks for the O4 pentasaccharide RU [5 anomeric, ring, two N-acetyl and two methyl (Rha & FucNAc)] groups as

well as EPA amino acids in characteristic regions. Each of the proton/carbon pairs for the O4 could be assigned based on the proton assignments and literature.

[00279] The structural characterization experiments thus confirmed that Glc-branched O4 bioconjugates (comprising polysaccharide antigen structures as indicated by Formula (O4-Glc+) in Table 1) could be produced, using the putative *E. coli* O4-gtrS gene identified in Example 2.

Example 4: Immunogenicity of a Glc-Branched O4 Bioconjugate in Rabbits

[00280] Glc-modified O4 bioconjugates (i.e. having glycans with the structure of Formula (O4-Glc+) as shown in Table 1) were used for rabbit immunization by applying a speedy-rabbit protocol (Eurogentec). Sera from immunized rabbits were analyzed by ELISA for anti-O4 IgG titers against purified O4 lipopolysaccharide (LPS) with (Glc+; i.e. containing glucosylated O4 polysaccharide) or without Glc-branching (Glc-; i.e. containing non-glucosylated O4 polysaccharide). Immunization with the bioconjugate resulted in high IgG titers in both rabbits (FIG. 1). In both cases, antibody titers induced by the O4 bioconjugate were higher against Glc+ LPS as compared to Glc- LPS.

[00281] Sera were also pooled and used in whole cell ELISA studies with test sets of *E. coli* O4 isolates with characterized gtrS status. Five gtrS-negative (no Glc-branching) and six gtrS-positive (Glc-branching) *E. coli* O4 isolates and a negative control strain were tested. Pooled sera from rabbits immunized with a Glc-modified O4 bioconjugate contained high titers of IgG specifically recognizing the tested O4 isolates (FIG. 2). In concordance with the LPS ELISA, all tested O4 isolates were recognized by the immune sera. The gtrS-positive isolates displayed an overall higher binding than the gtrS-negative isolates (FIG. 2). In particular, the following isolates were gtrS-positive: Y1382, E551, OC24334, stGVXN4983, stGVXN4994 and OC24794, and the following isolates were gtrS-negative: A2625, stGVXN4988, OC24784, OC24787, and OC24788. Immune sera did not bind the negative control strain of a non-related O-serotype, *E. coli* OC9487 (ATCC 35383).

[00282] The profiles of LPS extracted from the test set of gtrS-positive and -negative isolates in silver-stained polyacrylamide gels did not reveal marked differences between isolates expressing unmodified and modified forms of the O4 antigen confirming that the observed

differences are not explained by quantitative differences in LPS expression levels (data not shown).

[00283] Western blots of extracted LPS using pooled immune sera were performed to assess recognition of O4 O-antigen by IgGs elicited in response to immunization with a Glc-modified O4 bioconjugate. Binding of both modified and unmodified O4 LPS by IgGs from modified O4 immunized rabbits was observed and included specific recognition of LPS bands spanning a wide range of sizes, including high molecular weight LPS bands (FIG. 3).

[00284] In the further experiments below, when reference is made to ‘O4’ bioconjugate or production strains or ‘EcoO4’, the bioconjugate or production strain of Glc-branched O4 (having glycan structure (O4-Glc+) in Table 1) is meant, unless specifically indicated otherwise (the terms ‘O4’ and ‘O4-Glc+’ are thus used interchangeably for bioconjugates or production strains in those experiments).

Example 5: Immunogenicity of a Glc-Branched O4 Bioconjugate in Rats

[00285] Sprague Dawley rats were immunized intramuscularly 3 times with formulation buffer or (O4-Glc+)-EPA bioconjugate (i.e. bioconjugate of glucosylated O4 antigen polysaccharide covalently coupled to EPA carrier protein; carrier protein was EPA-2 as described in Example 3 above) at 3 different doses (0.04 µg, 0.40 µg or 4.0 µg). Serum antibody levels were measured by ELISA at day 0, 14 and 42 post-immunization.

[00286] Immunization with 0.04 µg, 0.40 µg and 4.00 µg of (O4-Glc+)-EPA bioconjugate induced significant increase in the levels of IgG antibodies at day 42 post-immunization when compared to formulation buffer (FIG. 4A). The antibodies induced by (O4-Glc+)-conjugate were functional, i.e., capable of mediating killing of (O4-Glc+) *E. coli* strain (FIG. 4B).

[00287] Antibody levels induced by 0.04 µg, 0.40 µg and 4.0 µg of (O4-Glc+)-EPA bioconjugate were significantly increased at day 42 as compared to those detected at baseline (day 42 vs day 0, P = 0.006 for all doses) and at day 14 post-immunization (day 42 vs day 14, P = 0.006 for all doses) (FIG. 5). In the group that received 4.0 µg of bioconjugate, titers were also significantly increased at day 14 compared to day 0, indicating that a single dose of 4.0 µg of (O4-Glc+)-EPA bioconjugate induces significant increase in IgG titers (day 14 vs day 0, P=0.012). The significant increase in IgG titers observed between day 14 and 42, for all three

concentrations of bioconjugate tested showed that a third dose of (O4-Glc+)-EPA bioconjugate is able to boost antibody responses (FIG. 5).

[00288] Functionality of antibodies induced by O4-Glc+-EPA conjugate in the rats immunized intramuscularly 3 times with formulation buffer or the bioconjugate at 4.00 µg/dose was determined by opsonophagocytic killing assay (OPKA) using O4(Glu+) and O4(Glu-) *E. coli* strains. The antibodies induced by (O4-Glc+)-EPA bioconjugate were functional, i.e., capable of mediating killing of an (O4-Glc+) *E. coli* strain (FIG. 4B, FIG. 6). Notably, antibodies induced by (O4-Glc+)-EPA bioconjugate were capable of mediating killing of both (O4-Glc+) and (O4-Glc-, i.e. having glycans with structure of Formula (O4-Glc-) in Table 1, i.e. O4 polysaccharide without Glc-branching) *E. coli* strains (FIG. 6).

[00289] In conclusion, antibodies induced by O4-Glc+-EPA bioconjugate are cross-reactive and capable of mediating killing of *E. coli* O4 strains with and without glucose branching.

Example 6: Production Strains for *E. coli* O-antigen Bioconjugates and resulting Bioconjugate products

[00290] In addition to (O4-Glc+)-EPA bioconjugates prepared as described above, nine (9) other bioconjugates were produced. In particular, the additionally produced bioconjugates included *E. coli* O1A-EPA bioconjugate, O2-EPA bioconjugate, O6A-EPA bioconjugate, O8-EPA bioconjugate, O15-EPA bioconjugate, O16-EPA bioconjugate, O18A-EPA bioconjugate, O25B-EPA bioconjugate, and O75-EPA bioconjugate. The chemical structures of the glycans of these conjugates can be seen in the respective Formulas in Table 1. A composition comprising the 10 bioconjugates is referred to herein as ‘ExPEC10V’. A composition comprising the O1A-EPA, O2-EPA, O6A-EPA and O25B-EPA bioconjugates is referred to as ‘ExPEC4V’ (and was previously described in for instance WO 2015/124769 and WO 2017/035181).

[00291] *Escherichia coli* W3110 Parental Strain

[00292] The non-pathogenic *E. coli* K12 strain W3110 was used as the parental strain for the construction of all ten production strains. The *E. coli* K12 strain W3110 was obtained from the Coli Genetic Stock Center (Yale University, New Haven (CT), USA, product number CGSC#4474). Its relevant genotype was previously described (*E. coli* W3110, F-, lambda-, IN(rrnD-rrnE)1, rph-1) and its genomic sequence was previously published (Hayashi K, et al.,

2006, Mol. Syst. Biol. 2006.0007 (doi:10.1038/msb4100049). The *E. coli* W3110 strain was genetically modified to enable production of each of the *E. coli* O-antigen bioconjugates (Table 3).

[00293] Bioconjugate production strains

[00294] The “ExPEC4V” and “ExPEC10V” compositions both comprise the O2-EPA and O25B-EPA bioconjugates from the same production strains. The “ExPEC4V” composition comprises the O1A-EPA bioconjugate from the stGVXN4411 or stLMTB10217 production strains, while the “ExPEC10V” composition comprises the O1A-EPA bioconjugate from the stLMTB10217 production strain. The “ExPEC4V” composition comprises the O6A-EPA bioconjugate from the stGVXN4112 production strain, while the “ExPEC10V” composition comprises the O6A-EPA bioconjugate from the stLMTB10923 production strain. Furthermore, the “ExPEC10V” composition comprises the O4-EPA (i.e. (O4-Glc+)-EPA), O8-EPA, O15-EPA, O16-EPA, O18A-EPA, and O75-EPA bioconjugates from production strains that are not used for “ExPEC4V”. Different production strains could vary in the plasmids for expression of the EPA carrier protein and/or the oligosaccharyl transferase PglB, as indicated below. An overview of several production strains is given in Table 3 below.

Table 3: Overview of genetic engineering of *E. coli* production strains for O-antigen bioconjugates for ExPEC4V and ExPEC10V vaccine compositions

Serotype	Strain name	Genomic mutations			Plasmids	
		<i>rfb</i> gene cluster	<i>waaL</i>	<i>gtrABS</i>	<i>pglB</i>	<i>epa</i>
O1A (ExPEC4V)	stGVXN4411	Δ <i>rfb</i> ::O1A <i>rfb</i> upcGVXN_032	Δ <i>waaL</i>	-	pGVXN970	pGVXN1076
O1A (ExPEC4V; ExPEC10V)	stLMTB10217	Δ <i>rfb</i> ::O1A <i>rfb</i> upcGVXN_032	Δ <i>waaL</i>	-	pGVXN1221	pGVXN1076
O2	stGVXN4906	Δ <i>rfb</i> ::O2 <i>rfb</i> upcGVXN_116	Δ <i>waaL</i>	-	pGVXN971	pGVXN1076
O4	BVEC-L-00684	Δ <i>rfb</i> ::O4 <i>rfb</i> CCUG11450	Δ <i>waaL</i>	Δ <i>gtrS::gtrS</i> O4	pGVXN1217	pGVXN1076

O6A (ExPEC4V)	stGVXN4112	$\Delta rfb::O6A\ rfb$ CCUG11309	$\Delta waal$	-	pGVXN114	pGVXN659
O6A (ExPEC10V)	stLMTB10923	$\Delta rfb::O6A\ rfb$ CCUG11309	$\Delta waal$	-	pGVXN1221	pGVXN1076
O8	stLMTB11734	$\Delta rfb::O8\ rfb$ E2420	$\Delta waal$	$\Delta gtrABS$	pGVXN970	pGVXN1076
O15	stLMTB11738	$\Delta rfb::O15\ rfb$ OC24891	$\Delta waal$	$\Delta gtrABS$	pGVXN1221	pGVXN1076
O16	stLMTB11739	$\Delta rfb::O16\ rfb$ OC24208	$\Delta waal$	$\Delta gtrABS$	pGVXN2381	pGVXN1076
O18A	BVEC-L-00559	$\Delta rfb::O18A\ rfb$ OC24255	$\Delta waal$	$\Delta gtrABS$	pGVXN970	pGVXN1076
O25B	stGVXN4459	$\Delta rfb::O25B\ rfb$ upecGVXN_138	$\Delta waal$	$\Delta gtrABS$	pGVXN970	pGVXN1076
O75	stLMTB11737	$\Delta rfb::O75\ rfb$ CCUG31	$\Delta waal$	$\Delta gtrABS$	pGVXN1217	pGVXN1076

[00295] O-antigen Biosynthesis (*rfb*) Gene Cluster

[00296] In all *E. coli* O-antigen production strains, the naturally occurring *E. coli* W3110 genomic O16::IS5 -antigen biosynthesis (*rfb*) gene cluster was replaced by the selected O-antigen-specific biosynthesis clusters from *E. coli* strains of the selected serotype, encoding for the serotype-specific O-antigen structures (see Table 1 for these O-antigen structures). The ten donor *rfb* clusters were selected or confirmed after whole-genome analysis of *E. coli* blood isolates. Replacement of the W3110 O16::IS5 *rfb* gene cluster, which is defective in O-antigen biosynthesis, has been achieved in a single homologous recombination event. In case of the O16 and O18A *rfb* gene clusters, the donor DNA recombined via the flanking *gnd* and *rmlCA* genes, while the *rfb* gene cluster for the other strains recombined via the flanking *gnd* and *galF* genes. Sequences of the *rfb* clusters in the production strains are provided in SEQ ID NOS: 9 and 11-19.

[00297] O-antigen ligase (*waaL*) gene

[00298] All *E. coli* O-antigen production strains carry an artificially introduced deletion of the *E. coli* W3110 genomic O-antigen ligase encoded by the *waaL* gene. In the $\Delta waaL$ strains the transfer of the O-antigen to lipid A is disrupted, which instead directs transfer of the O-antigen to the carrier protein to increase product yield.

[00299] O-antigen glucosylation (*gtrABS*) genes

[00300] In the *E. coli* O8, O15, O16, O18A, O25B, and O75 production strains the *E. coli* W3110 genomic *gtrABS* genes, which are responsible for O16 O-antigen glucosylation, have been deleted. While the *gtrA* and *gtrB* genes in different serotypes are highly homologous and interchangeable, the *gtrS* gene encodes a serotype-specific O-antigen glycosyl transferase. In *E. coli* W3110 GtrS can transfer a glucose (Glc) residue to the GlcNAc sugar in the α -L-Rha-(1 \rightarrow 3)-D-GlcNAc motif of the *E. coli* O16 O-antigen. In the *E. coli* O1A, O2 and O6A production strains no deletion or replacement of the *gtrABS* gene has occurred. These O-antigens miss the α -L-Rha-(1 \rightarrow 3)-D-GlcNAc motif that is the natural substrate for *E. coli* O16 *gtrS*. In the *E. coli* O4 production strain, the W3110 *gtrS* gene has been replaced with the *E. coli* O4 *gtrS* gene to accommodate proper glucosylation of the *E. coli* O4 O-antigen.

[00301] Oligosaccharyl transferase PglB

[00302] All *E. coli* O-antigen production strains expressed a variant of the *C. jejuni* glycosyl transferase PglB, which can transfer the O-antigen onto an amino acid consensus sequence on a carrier protein by *N*-glycosylation. PglB has broad substrate recognition, but due to low product yields several production strains were prepared expressing a PglB variant having modified substrate specificities, which resulted in improved product yield (see e.g. WO 2016/107818, WO 2016/107819). The *pglB* gene was placed behind an Isopropyl β -D-1-thiogalactopyranoside (IPTG) inducible promoter on a plasmid. Table 4 below lists the PglB variants encoded by the plasmids used for production of the *E. coli* O-antigen production strains for the bioconjugates for the ExPEC4V and ExPEC10V compositions described above. Further plasmids with variation in vector backbone, antibiotic resistance marker, and/or alternative PglB variants have also been tested successfully for bioconjugate production.

[00303] **Table 4:** PglB and EPA plasmids used in *E. coli* O-antigen Production Strains

Plasmid name	Gene	Description ¹
pGVXN114	<i>pglB</i>	<i>C. jejuni</i> codon usage; SpR
pGVXN970	<i>pglB</i>	<i>E. coli</i> codon usage optimized; SpR
pGVXN971	<i>pglB^{N534Q}</i>	<i>E. coli</i> codon usage optimized; The natural glycosylation site of PglB was inactivated; SpR
pGVXN1217	<i>pglB^{N311V}</i>	<i>E. coli</i> codon usage optimized; Substrate optimized PglB; SpR
pGVXN1221	<i>pglB^{N311V,K482R,D483H,A669V}</i>	<i>E. coli</i> codon usage optimized; Substrate optimized PglB; SpR
pGVXN2381	<i>pglB^{Y77H,S80R,Q287P,K289R,N311V}</i>	<i>E. coli</i> codon usage optimized; Substrate optimized PglB; SpR
pGVXN659	EPA-4	EPA with four bioconjugation sites; AmpR
pGVXN1076	EPA-4	EPA with four bioconjugation sites; KanR

¹ SpR, spectinomycin resistant; AmpR, ampicillin resistant; KanR, kanamycin resistant

[00304] Carrier protein (EPA)

[00305] All *E. coli* O-antigen production strains expressed a genetically detoxified *P. aeruginosa* ADP-ribosyltransferase toxoid (EPA) as a carrier protein for the O-antigen. The EPA toxoid differs from wild-type EPA toxin in two residues: Leu552 was changed to Val and Glu553 (in the catalytic domain) was deleted. Glu553 deletions were reported to significantly reduce toxicity. In addition to the detoxification mutation, four (EPA-4) consensus *N*-glycosylation site motifs were introduced. The *epa* gene was placed behind a L-Arabinose (Ara) inducible promoter on a plasmid (Table 4). Table 4 is limited to the plasmids used in production strains for bioconjugates used in the “ExPEC4V” and “ExPEC10V” compositions described above. Plasmids with variation in vector backbone, antibiotic resistance marker, and/or EPA variants, e.g. varying in the number of consensus *N*-glycosylation site motifs (e.g. having two such motifs, EPA-2), have also been tested successfully for bioconjugate production.

Example 7: Optimizing the oligosaccharyltransferase for generation of bioconjugates with glucosylated O4 (O4-Glc+) antigen

[00306] Yield optimization for bioconjugate production can be achieved by modification of the *C. jejuni* oligosaccharyl transferase PglB, which can lead to a more efficient or higher degree of N-glycosylation of the O-antigen of interest to the EPA carrier protein. In an *E. coli* strain for production of bioconjugate with glucosylated O4 (O4-Glc+) O-antigen polysaccharide, such optimization strategy was applied and resulted in an (O4-Glc+)-specific optimized PglB variant improving bioconjugate product yield.

[00307] In this approach, an O4-Glc+ O-antigen polysaccharide producing strain containing an EPA-expression plasmid was transformed with a variety of different PglB expression plasmids, each of which contained different amino acid substitutions in the PglB protein, altering substrate specificity. Bioconjugate production level and profile of each strain was assessed at shake-flask level in osmotic shock experiments, and readout was performed by capillary electrophoresis immunoassays on the periplasmic extract using O4-Glc+ -specific monoclonal antibodies.

[00308] One of the tested PglB variants containing an N311V amino acid substitution was found to improve product yield of glucosylated O4 bioconjugates significantly (FIG. 7A).

[00309] In a further improvement where the N311V PglB-variant was further modified, an Y77H amino acid substitution further enhanced O4-Glc+-specific product yield and showed an increased degree of di-and tri-glycosylated product compared to the N311V PglB-variant, where other modifications were found to be neutral or had a negative effect on product yield (FIG. 7B). Plasmid pLMTB4008 (SpR) encodes *E. coli* codon usage optimized, (O4-Glc+)-substrate optimized, PglB variant with mutations Y77H and N311V.

[00310] The PglB variant with optimized substrate specificity for O4-Glc+ O-antigen polysaccharide, containing N311V and Y77H amino acid substitutions relative to wild-type (wt) *C. jejuni* glycosyl transferase PglB, was found to double bioconjugate yield compared to the first round optimized PglB-N311V variant.

[00311] Similarly using screens, the most optimal yielding PglB variants were also determined for *E. coli* O-antigen bioconjugate production of the of the other nine serotypes in the ExPEC10V composition.

[00312] For bioconjugates having the O1A, O6A, or O15 antigen polysaccharide, PglB with amino acid mutations N311V, K482R, D483H, and A669V was found to give the highest yields.

[00313] For bioconjugates having the O2, O8, O18A, or O25B antigen polysaccharide, wild-type PglB (i.e. not having amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483 and 669) was found to give the highest yields.

[00314] For bioconjugates having the O16 antigen polysaccharide, PglB with amino acid mutations Y77H, S80R, Q287P, K289R, and N311V was found to give the highest yields.

[00315] For bioconjugates having the O75 antigen polysaccharide, PglB with amino acid mutation N311V was found to give the highest yields.

[00316] It can be seen from these results that the optimal PglB variant is different for different O-antigens, and that the optimal PglB variant for producing a bioconjugate with a given O-antigen polysaccharide is unpredictable.

Example 8: Bioconjugates of O-antigens from 10 *E. coli* serotypes and their quality attributes

[00317] O-glycan residues of the target O-antigens are structurally diverse and have variable repeating units. The specificity and affinity of the glycosyl transferase PglB is linked to the glycan structure. Thus, making a bioconjugate that has the desired quality attributes, e.g., purity, glycan/protein ratio, etc., is a challenging, non-straightforward, task. The right combination of PglB and EPA carrier protein determines the yield and may influence glycosylation efficiency. By optimizing the PglB and carrier proteins, bioconjugates having the desired quality attributes were produced. It may be also important to maintain a lower threshold value of total carrier protein, particularly when one or more O-antigen bioconjugates are combined together and administered in a single composition or vaccine, because very high amounts of carrier protein may lead to immunological interference. In order to avoid such a phenomenon, conjugates having a higher glycan/protein ratio are preferred. Hence, for ExPEC10V vaccine, bioconjugates with at least comparable (to the previously described ExPEC4V vaccine that has been subject to clinical trials) glycosylation ratio were developed.

[00318] The bioconjugates were each produced by culturing the respective host cells (Example 6, Table 3) in bioreactors (10L and/or 200L volumes) and expression of the bioconjugates, following methods previously described. Each drug substance was manufactured batch-wise by bacterial fed-batch fermentation to generate biomass containing the expressed bioconjugates of

the corresponding polysaccharide serotype. Cells were cultured and induced with IPTG and arabinose. The bioconjugates were isolated from the periplasm of the cells in the bioreactor cultures by osmotic shock followed by chromatographic purification. This process was performed for each of the 10 bioconjugates.

[00319] The *E. coli* O-antigen bioconjugates thus prepared that are drug substances (DSs) for ExPEC10V and ExPEC4V showed comparable critical quality attributes: (1) process-related purity (measured by RP-HPLC) was higher than 95%, (2) polysaccharide/protein ratio ranged between about 0.1-0.5, mostly between 0.15 and 0.45, (3) bacterial endotoxin (Ph. Eur. 2.2.3) was less than 0.5 EU/ μ g polysaccharide. The average length of the individual polysaccharide chains was typically between about 10-20 repeating units (measured using high resolution SDS-PAGE).

[00320] The structures of the polysaccharide repeat units were confirmed (by NMR and MS/MS of the conjugates, intact or trypsin-digested) to be the ones shown in the Formulas for the corresponding serotypes in Table 1, for all ten bioconjugates that are DSs for the ExPEC10V composition described above.

[00321] The O18 serotype had the lowest yields of bioconjugate production amongst the ten serotypes of which bioconjugates were made for the ExPEC10V composition.

[00322] ExPEC10V drug product (DP) comprises a mixture of the ten monovalent DSs described above.

Example 9: Toxicology of ExPEC10V vaccine

[00323] A single-dose pilot toxicity and local tolerance study (non-GLP) with ExPEC10V was conducted in female NZW rabbits. One group (n=2) received an intramuscular (IM) injection (on Day 0) of the control (saline), and a second group (n=4) received an IM injection of ExPEC10V at 105.6 μ g total polysaccharide (PS)/dose (9.6: 9.6: 9.6: 9.6: 9.6: 9.6: 9.6: 9.6: 19.2: 9.6 μ g PS per dose, for respectively O-serotypes O1A, O2, O4, O6A, O8, O15, O16, O18A, O25B and O75) using a dosing volume of 0.6 mL (176 μ g PS/mL). Necropsy was performed on Day 2.

[00324] There were no mortalities observed. In addition, there were no vaccine-related effects noted for clinical observations (including injection site effects using Draize scoring), body weight, food consumption, and body temperature. Histopathologically, there were no vaccine-

related changes observed at the administration site or draining (iliac) lymph node. A minimal increase in germinal center formation in the spleen was observed in one out of four treated animals (Day 2), and was considered a normal, immunological response to the injected vaccine. Overall, the administration of a single IM dose of ExPEC10V to female rabbits was well-tolerated.

Example 10: Immunogenicity of ExPEC10V blended formulation in rabbits

[00325] An ExPEC4V vaccine (comprising bioconjugates of *E. coli* O1A, O2, O6A, and O25B serotypes) has previously been shown to be immunogenic for these four serotypes in rats, rabbits, and humans (see e.g. WO 2015/124769; WO 2017/035181; Huttner et al, 2017, Lancet Infect Dis, [http://dx.doi.org/10.1016/S1473-3099\(17\)30108-1](http://dx.doi.org/10.1016/S1473-3099(17)30108-1); RW Frenck Jr, et al, abstract 5587, ASM Microbe 2018). The novel bioconjugates of the invention having the *E. coli* glucosylated O4 serotype were shown to be immunogenic in Examples 4 and 5 above. Immunogenicity of the bioconjugates of *E. coli* serotypes O8, O15, O16, O18A, and O75 (all having EPA-2 as carrier protein in this experiment) when separately administered (monovalent) to rats confirmed that also each of these bioconjugates was immunogenic, since ELISA data indicated that each of these bioconjugates could elicit high levels of *E. coli* O-antigen specific antibodies (not shown).

[00326] Immunogenicity of the 10-valent vaccine that contained a mixture of the 10 bioconjugates as described above was also tested. New Zealand White (NZW) rabbits (female, 12-16 weeks old) received 3 intramuscular immunizations with ExPEC10V or saline administered 2 weeks apart (Table 5; administration at days 0, 14, and 27). The 10 polysaccharides that are part of the ExPEC10V vaccine used in these experiments were conjugated to the carrier protein EPA containing 4 sites of glycosylation (EPA-4). The vaccine was formulated in 3 different doses: Group 1 ('high dose'): 8 ug/dose of O1A, O2, O6A, O4, O8, O15, O16, O18 and O75 and 16 ug/dose of O25B; Group 2 ('medium dose'): 4 ug/dose of O2, O4, O8, O15, O16, O18 and O75, 8 ug/dose of O1A and O6A and 16 ug/dose of O25B; Group 3 ('low dose'): 0.4 ug/dose of O2, O4, O8, O15, O16, O18 and O75, 0.8 ug/dose of O1A and O6A and 1.6 ug/dose of O25B. Animals from the control group (Group 4) received only saline (0.9% (w/v) sodium chloride solution) (Table 5).

[00327] Antibody responses were evaluated at day 0 (pre-immunization) and days 14, 27 and 42 post-immunization. Serum antibody levels induced by each of the bioconjugates included in the vaccine and the carrier protein EPA were measured by ELISA (total IgG), using type-specific LPS as coating material. The antibody titers were reported as EC50 values that correspond to the half maximal effective concentration based on duplicates of 12-step titration curves plotted in a 4-parameter logistic nonlinear regression model. Functional activity was determined by OPK.

Table 5. Description of experimental groups.

Experimental groups	Dosing ($\mu\text{g/PS}$) O1A:O2:O6A:O25B:O4:O8:O15:O16:O18A:O75	Sample size
Group 1 (high dose)	8:8:8:16:8:8:8:8:8:8	7
Group 2 (medium dose)	8:4:8:16:4:4:4:4:4:4	7
Group 3 (low dose)	0.8:0.4:0.8:1.6:0.4:0.4:0.4:0.4:0.4:0.4	7
Group 4 (control)	0.9% (w/v) sodium chloride solution	7

[00328] Results are shown in FIG. 8 and summarized in Table 6.

[00329] **Table 6.** Summary of *E. coli* O-antigen specific antibody responses induced by ExPEC10V in NZW rabbits.

ExPEC10V		Antibody responses day 0 post-vaccination									
dose	O1A	O2	O6A	O25B	O4	O8	O15 [#]	O16	O18A	O75	
High	++	++	++	++	++	++	++	++	++	++	++
Mid	++	++	++	++	++	++	++	++	++	++	++
Low	++	++	++	++	++	++	++	++	++	++	++

ExPEC10V		Antibody responses day 27 post-vaccination									
dose	O1A	O2	O6A	O25B	O4	O8	O15 [#]	O16	O18A	O75	
High	++	++	++	++	++	++	++	++	++	++	++
Mid	++	++	++	++	++	++	++	++	++	++	++
Low	++	++	++	++	++	++	++	++	++	++	++

ExPEC10V		Antibody responses day 42 post-vaccination									
dose	O1A	O2	O6A	O25B	O4	O8	O15 [#]	O16	O18A	O75	
High	++	++	++	++	++	++	++	++	++	++	++
Mid	++	++	++	++	++	++	++	++	++	++	++
Low	++	++	++	++	++	++	++	++	++	++	++

Dark gray squares show serotype-specific antibody responses in which p values were statistically significant. Light gray squares show serotype-specific antibody responses in which p values were not statistically significant (ns). Wilcoxon Rank Sum test with Bonferroni correction for multiple comparisons. Comparisons ExPEC10V vaccinated animals (Group 1, 2 and 3) versus saline control (Group 4). * $p\leq 0.05$, ** $p\leq 0.01$. # P values were statistically significant after excluding an outlier animal from the control group (sensitivity analysis).

[00330] The high dose of ExPEC10V (Group 1) induced significantly higher IgG antibody levels at all time-points investigated (Days 14, 27 and 42 post-immunization) when compared to saline control for O1A, O2, O4, O6A, O16, O18A and O25B (FIG. 8, Table 6). Significantly higher antibody titers induced by O8 and O75 conjugates when compared to saline control were observed at Days 27 and 42 post-immunization (FIG. 8, Table 6).

[00331] The medium dose of ExPEC10V (Group 2) and the low dose (Group 3) induced significantly higher antibody levels at all time-points investigated (Days 14, 27 and 42 post-immunization) when compared to saline control for O1A, O2, O4, O6A, O16 and O25B (FIG. 8, Table 6). Significantly higher antibody titers induced by O8, O18A and O75 conjugates when compared to saline control were observed at Days 27 and 42 post-immunization suggesting that the boost dose in rabbits increases the response to these O-serotypes (FIG. 8, Table 6).

[00332] For O15 conjugates, sensitivity analysis omitting an outlier animal from the control group showed that all three doses of ExPEC10V vaccine induced a significant increase in antibody responses when compared to saline control at Days 14, 27 and 42 post-immunization (FIG. 8, Table 6).

[00333] Antibodies induced by the carrier protein EPA were significantly higher than EPA antibody titers in the saline-treated (control) group for the three doses of ExPEC10V tested (high, medium and low) at all time points investigated (Days 14, 27 and 42) (FIG. 8).

[00334] Between dose comparisons (not shown) showed that at Day 14 post-vaccination, the high dose of ExPEC10V induced significantly higher antibody responses when compared to the low dose for most of the conjugates tested (O1A, O2, O4, O6A, O15, O16, O18A and O25B). The medium dose of ExPEC10V also induced significantly higher antibody responses compared

to the low dose for O1A, O2, O4, O18A, O25B and O75. For O8 conjugate, all three formulations of ExPEC10V induced similar levels of antibodies at Day 14 post-vaccination.

[00335] The low dose of ExPEC10V induced a significant increase in antibody responses at Day 42 post vaccination (after a prime and two boost doses) when compared to the high and medium doses of ExPEC10V for O1A, O2, O4, O16, O25B and O75 conjugates. These findings are in line with other experiences with conjugate vaccines, where for instance no clear relationship between dose and the magnitude of the antibody response to primary vaccination was observed in infants vaccinated with pneumococcal conjugate vaccine (Poolman JT, et al. Expert Rev Vaccines. 2013, 12(12):1379-94).

[00336] There were no significant differences between the three doses of ExPEC10V tested at Day 42 post-vaccination for O6A, O8 and O15 conjugates. For the O18A conjugate, the high dose of ExPEC10V induced a significantly higher antibody response when compared to the medium dose at Day 42 post-vaccination.

[00337] For the carrier protein (EPA), the high and medium dose of ExPEC10V induced significantly higher antibody responses when compared to the low dose at day 14 post-vaccination. The high dose of the vaccine also induced significantly higher antibody responses when compared to the low dose at day 42 post-vaccination.

[00338] In conclusion, the three formulations of ExPEC10V (high, medium and low), administered via intramuscular injection on Days 0, 14, 27 are immunogenic in rabbits.

[00339] So far, functional antibodies capable of killing *E.coli* strains induced by this vaccine in rabbits were shown for serotypes O1A, O2, O4, O6A, O15, O16 and O25B.

[00340] In a further experiment, a GMP batch of the ExPEC10V vaccine (see Example 8 above for production) was prepared and injected into NZW rabbits as part of a toxicology study (Table 7). In this study, NZW rabbits (males and females) received 3 intramuscular injections (0.6 mL) of the ExPEC10V vaccine (day 1, 15 and 29) and a control group received 0.9% (w/v) sodium chloride solution (saline). Each dose of the vaccine contained 9.6 µg polysaccharide (PS) for serotypes O1A, O2, O4, O6A, O8, O15, O16, O18A and O75 and 19.2 µg PS for serotypes O25B, corresponding to 105.6 µg total PS (176 µg total PS/mL) and 382.8 µg of total EPA (638 µg EPA/mL). IgG titers against O-antigens and carrier protein (EPA) were determined from samples collected during the pre-treatment period (day 1) and days 31 and 50 post-immunization.

[00341] A significant increase in antibody responses against all O-antigens and the carrier protein EPA were observed at day 31 and 50 post-vaccination in the group that received ExPEC10V when compared to the control group that received only saline (Fig. 9, Table 8). For O1A serotype, a significantly higher antibody response was also observed at day 1 (baseline) when vaccinated animals were compared with the controls. These results suggest that some animals were pre-exposed to *E. coli* or have antibodies that cross-react with O1A-LPS.

[00342] Table 7. Experimental groups and ExPEC10V dose used in NZW rabbits.

Groups	Treatment	Dose	Dosing days	Main (day 31) (males/females)	Recovery (day 50) (males/females)
1	control	0	1, 15, 29	10	10
2	ExPEC10V	105.6 µg PS*	1, 15, 29	10	10

*Each dose (0.6 mL dosing volume) contains 9.6:9.6:9.6:9.6:9.6:9.6:9.6:9.6:19.2:9.6 µg polysaccharide (PS) for serotypes O1A, O2, O4, O6A, O8, O15, O16, O18A, O25B, O75, respectively (176 µg total PS/mL). Each dose contains 382.8 µg EPA protein (638 µg EPA/mL).

Table 8. Immunogenicity of ExPEC10V in NZW rabbits as part to a toxicology study.

Antibody responses Significant Serotype	Antibody responses (day 31 post-vaccination)									
	O1A	O2	O4	O6A	O8	O15	O16	O18A	O25B	O75
O1A	Significant									
O2		Significant								

Antibody responses induced by ExPEC10V. Light gray squares show serotypes in which a significant increase in antibody responses was observed in the vaccine group compared to control. Tobit model with a likelihood ratio test.
****P≤0.0001.

Example 11: Phase 1/2a trial with the ExPEC10V vaccine in humans

[00343] At present, there is no vaccine available to prevent IED. The serotypes comprising the ExPEC10V vaccine (O1A, O2, O4, O6A, O8, O15, O16, O18A, O25B and O75) were selected to address invasive disease caused by the majority of clinically relevant ExPEC strains that also represent the majority of ExPEC isolates causing antimicrobial resistant IED, including ST131. The selected serotypes are representative for the ten prevalent ExPEC O-serotypes causing

bloodstream infections in the older population and responsible for approximately 70% of bloodstream infections caused by ExPEC.

[00344] Since the mechanism of action of conjugate vaccines in the prevention of invasive disease is not expected to be affected by antibiotic resistance mechanisms, it is believed that ExPEC10V vaccine provides protection against IED caused by drug-resistant- and drug-susceptible O1A, O2, O4, O6A, O8, O15, O16, O18A, O25B and O75 serotypes.

[00345] There is preceding clinical experience with ExPEC4V, an earlier vaccine candidate which comprised a subset of four of the *E. coli* O-antigen conjugates (O1A, O2, O6A and O25B) also found in ExPEC10V. Based on the results from four clinical studies (two completed phase 1 studies, one completed phase 2 study and an ongoing phase 2 study), ExPEC4V was well-tolerated by the study participants and no vaccine-related safety signals were observed at doses up to 16 µg polysaccharide (PS) per serotype (O1A, O2, O6A and O25B). Most adverse events (AEs) were Grade 1 and 2, very few Grade 3 AEs were reported. Late-onset solicited local AEs (AEs which start after Day 5 post-vaccination) were observed mainly with the higher doses of ExPEC4V. In each study, the ExPEC4V vaccine was shown to be immunogenic, demonstrating a dose-dependent vaccine immune response, and O-antigen specific Immunoglobulin G (IgG) titer increases, as measured by enzyme-linked immunosorbent assay (ELISA). Functional activity of the antibodies was demonstrated with an ExPEC4V-optimized opsonophagocytic killing assay (OPKA). Co-analysis of ELISA and OPKA test results showed correlation between the assay responses (Pearson correlation coefficients ≥0.61 and ≥0.48 for Day 30 and Day 360, respectively in a Phase 2 clinical trial [study 4V-BAC2001]), substantiating the use of ELISA as a primary measure of ExPEC4V antibody titers and to predict functional antibody activity. Analysis of the immunogenicity data has demonstrated the durability of the immune response through three years after vaccination with ExPEC4V. It has now also been observed that sera from humans vaccinated with ExPEC4V and that had high titers of serotype-specific opsonophagocytic antibodies, when passively transferred into mice that were subsequently intraperitoneally challenged with *E. coli* strains of O25B or O2 serotype, were able to mediate protection *in vivo* (not shown). Hence, ExPEC4V-specific opsonophagocytic human antibodies mediate bacterial killing *in vivo*, which is in line with other conjugate vaccines

in which the proposed mechanism of protection is by induction of opsonophagocytic antibodies that mediate bacterial killing.

[00346] ExPEC10V includes a total of ten serotypes and increases coverage from about 50% (ExPEC4V) to approximately 70% of bloodstream infections caused by ExPEC in adults aged 60 years and older. Based on the clinical experience with ExPEC4V, and on the pre-clinical data for ExPEC10V as discussed in the examples above, it is expected that administration of ExPEC10V will induce immune responses to *E. coli* serotypes O1A, O2, O4, O6A, O8, O15, O16, O18A, O25B and O75 also in humans.

[00347] A randomized, observer-blind, first-in-human phase 1/2a study to evaluate the safety, reactogenicity, and immunogenicity of three different doses of the ExPEC10V vaccine is conducted in humans aged 60 to 85 years in stable health (study 10V-BAC1001). The study design includes 2 cohorts: A total of 1,004 participants are enrolled in the study with 404 participants (100 participants/ExPEC10V dose) aged ≥ 60 to ≤ 85 years in stable health in Cohort 1 and an additional of 600 participants aged ≥ 60 years in stable health with a history of UTI in the past 5 years in Cohort 2.

[00348] ExPEC10V is a 10-valent vaccine candidate in development for the prevention of invasive extraintestinal pathogenic *Escherichia coli* (ExPEC) disease (IED) in adults 60 years of age and older. ExPEC10V consists of the O-antigen polysaccharides (PSs) of the ExPEC serotypes O1A, O2, O4, O6A, O8, O15, O16, O18A, O25B and O75 separately bioconjugated to the carrier protein, a genetically detoxified form of exotoxin A (EPA) derived from *Pseudomonas aeruginosa*, and its production has been described above. The O4 PS is the glucosylated form, having the structure of Formula (O4-Glc+) in Table 1.

[00349] OBJECTIVES AND ENDPOINTS

[00350] COHORT 1 - Phase 1/2a observer-blind period with open-label long-term follow-up period (N=404):

Objectives	Endpoints
Primary	
<ul style="list-style-type: none"> To evaluate the safety and reactogenicity of different doses of ExPEC10V in participants ≥ 60 to ≤ 85 years of age 	<ul style="list-style-type: none"> Solicited local and systemic adverse events (AEs) collected for 14 days post-vaccination (from Day 1 to Day 15) Unsolicited AEs collected from the administration of the study vaccine until 29 days post-vaccination (from Day 1 to Day 30) Serious adverse events (SAEs) collected from the administration of the study vaccine until Day 181
<ul style="list-style-type: none"> To evaluate the dose-dependent immunogenicity of ExPEC10V on Day 15 in participants ≥ 60 to ≤ 85 years of age 	<ul style="list-style-type: none"> Antibody titers for ExPEC10V, as determined by multiplex electrochemiluminescent (ECL)-based immunoassay and multiplex opsonophagocytic assay (MOPA) on Day 15
Secondary	
<ul style="list-style-type: none"> To evaluate the correlation between multiplex ECL-based immunoassay (total antibody) and MOPA (functional antibody) serum titers on Day 15 	<ul style="list-style-type: none"> Antibody titers for ExPEC10V, as determined by multiplex ECL-based immunoassay and MOPA on Day 15
<ul style="list-style-type: none"> To evaluate the dose-dependent immunogenicity of ExPEC10V on Days 30 and 181 in participants ≥ 60 to ≤ 85 years of age 	<ul style="list-style-type: none"> Antibody titers for ExPEC10V, as determined by multiplex ECL-based immunoassay and MOPA on Days 30 and 181

Objectives	Endpoints
<ul style="list-style-type: none"> To evaluate, in the long-term follow-up (LTFU) period, the safety of the ExPEC10V dose selected for further clinical development based on the Day 30 primary analysis in participants ≥ 60 to ≤ 85 years of age 	<ul style="list-style-type: none"> SAEs related to the study vaccine or study procedures collected from Day 182 until the end of the study
<ul style="list-style-type: none"> To evaluate, in the LTFU period, the immunogenicity of the ExPEC10V dose selected for further clinical development based on the Day 30 primary analysis 	<ul style="list-style-type: none"> Antibody titers for ExPEC10V, as determined by multiplex ECL-based immunoassay and MOPA at Year 1 (Day 366), Year 2 (Day 731) and Year 3 (Day 1096)

COHORT 2 - Double-blind period with double-blind long-term follow-up period (N=600):

Objectives	Endpoints
Primary	
<ul style="list-style-type: none"> To evaluate the safety and reactogenicity of the selected dose of ExPEC10V in participants ≥ 60 years of age with a history of UTI in the past 5 years 	<ul style="list-style-type: none"> Solicited local and systemic AEs collected for 14 days post-vaccination (from Day 1 to Day 15) Unsolicited AEs collected from the administration of the study vaccine until 29 days post-vaccination (from Day 1 to Day 30) SAEs collected from the administration of the study vaccine until Day 181
<ul style="list-style-type: none"> To evaluate the immunogenicity of the selected dose of ExPEC10V on Day 30 in participants ≥ 60 years of age with a history of UTI in the past 5 years 	<ul style="list-style-type: none"> Antibody titers for ExPEC10V, as determined by multiplex ECL-based immunoassay and MOPA on Day 30
Secondary	
<ul style="list-style-type: none"> To evaluate the correlation between multiplex ECL-based immunoassay (total antibody) and MOPA (functional antibody) serum titers on Day 30 in participants ≥ 60 years of age with a history of UTI in the past 5 years 	<ul style="list-style-type: none"> Antibody titers for ExPEC10V, as determined by multiplex ECL-based immunoassay and MOPA on Day 30
<ul style="list-style-type: none"> To evaluate the immunogenicity of the selected dose of ExPEC10V on Days 15 and 181 in participants ≥ 60 years of age with a history of UTI in the past 5 years 	<ul style="list-style-type: none"> Antibody titers for ExPEC10V, as determined by multiplex ECL-based immunoassay and MOPA on Days 15 and 181

Objectives	Endpoints
<ul style="list-style-type: none"> To evaluate, in the LTFU period, the safety of the selected dose of ExPEC10V in participants ≥ 60 years of age with a history of UTI in the past 5 years 	<ul style="list-style-type: none"> SAEs related to the study vaccine or study procedures collected from Day 182 until the end of the study
<ul style="list-style-type: none"> To evaluate, in the LTFU period, the immunogenicity of the selected dose of ExPEC10V in participants ≥ 60 years of age with a history of UTI in the past 5 years 	<ul style="list-style-type: none"> Antibody titers for ExPEC10V, as determined by multiplex ECL-based immunoassay and MOPA at Year 1 (Day 366), Year 2 (Day 731), and Year 3 (Day 1096)
Exploratory	
<ul style="list-style-type: none"> To evaluate the effect of ExPEC10V on the intestinal (stool) microbiome by metagenomic analyses 	<ul style="list-style-type: none"> Metagenomics of stool samples from a selected subset¹ of participants to evaluate the effect of ExPEC10V on: <ul style="list-style-type: none"> Prevalence of pathogens (eg, <i>Clostridium difficile</i>) in the intestinal flora Prevalence of ExPEC10V serotypes in the intestinal flora

[00351] OVERALL DESIGN

[00352] This is a randomized, multicenter, interventional study including two cohorts.

[00353] For Cohort 1, the study has an observer-blind, active-controlled design, and a total of 404 adult participants aged ≥ 60 to ≤ 85 years in stable health with or without a history of UTI are included. The study design for Cohort 1 is comprised of three periods: a maximum of 28-day screening period, an observer-blinded 181-day follow-up period with vaccination on Day 1 and an open-label LTFU period which lasts from Day 182 until 3 years (Day 1096) post-vaccination (FIG. 10A). Only participants from the ExPEC10V selected dose group (approximately 100

participants) and participants from the Prevnar 13 group progress to the LTFU period. The end of Cohort 1 is the last participant's Year 3 visit (Day 1096).

[00354] For Cohort 2, the study has a double-blind, placebo-controlled design, and a total of 600 adult participants aged ≥ 60 years in stable health with a history of UTI in the past 5 years is included. Enrollment commences after completion of the Phase 1/2a primary analysis and ExPEC10V dose selection from Cohort 1. The study design for Cohort 2 is comprised of three periods: a maximum 28-day screening period, a double-blind 181-day follow-up period with vaccination on Day 1, and a double-blind LTFU period which lasts from Day 182 until 3 years (Day 1096) post-vaccination (Fig. 10B). All participants in Cohort 2 progress to the LTFU period. The end of study is the last participant's Year 3 visit (Day 1096) in Cohort 2.

[00355] Cohort 1: Phase 1

[00356] In Phase 1 of Cohort 1, a total of 84 participants are enrolled in a staggered approach following stepwise dose-escalating procedures with safety evaluations in place before progressing from one step to the next. An internal Data Review Committee (DRC) is commissioned for this study to review the physical examination data (baseline as well as targeted), baseline demographic data and the 14-day post-vaccination safety data (including solicited local and systemic AEs, unsolicited AEs, SAEs, clinical laboratory data and vital signs) of these 84 Phase 1 participants. In this phase of the study, participants were enrolled and randomized in six steps:

Step 1: Four sentinel participants were enrolled and randomized; two participants in the ExPEC10V low dose group (Table 11), and one participant each in the ExPEC4V and Prevnar 13 groups.

Step 2: Twenty-four participants were enrolled and randomized; 18 participants in the ExPEC10V low dose group (Table 11), and three participants each in the ExPEC4V and Prevnar 13 groups.

Step 3: Four sentinel participants were enrolled and randomized; two participants in the ExPEC10V medium dose group (Table 11), and one participant each in the ExPEC4V and Prevnar 13 groups.

Step 4: Twenty-four participants were enrolled and randomized; 18 participants in the ExPEC10V medium dose group (Table 11), and three participants each in the ExPEC4V and

Prevnar 13 groups.

Step 5: Four sentinel participants were enrolled and randomized; two participants in the ExPEC10V high dose group (Table 11), and one participant each in the ExPEC4V and Prevnar 13 groups.

Step 6: Twenty-four participants were enrolled and randomized; 18 participants in the ExPEC10V high dose group (Table 11), and three participants each in the ExPEC4V and Prevnar 13 groups.

[00357] All participants received a single intramuscular (IM) injection of either ExPEC10V (1 of 3 doses), ExPEC4V or Prevnar 13 on Day 1 per the assigned study vaccination groups. The four sentinel participants at each of Steps 1, 3 and 5 were contacted by telephone 24 hours post-vaccination to collect safety information. The blinded 24-hour post-vaccination safety data in each group of four sentinel participants were reviewed by the principal investigator (PI), study responsible physician (SRP) and sponsor medical lead (SML). Randomization of additional participants for the next step was halted until this Day 2 sentinel safety evaluation was completed.

[00358] In the absence of any clinically significant findings, an additional 24 participants (for Steps 2, 4, and 6) were enrolled and randomized to one of three study vaccination groups (Table 11) to receive a single IM injection of either ExPEC10V (1 of 3 doses), ExPEC4V or Prevnar 13 on Day 1.

[00359] After vaccination of an additional 24 participants at each dose level (low dose in Step 2, medium dose in Step 4, and high dose in Step 6), 14-day post-vaccination safety data of all 28 (4+24) participants at each dose level was reviewed by the DRC before progressing to the next dose level or Phase 2a.

[00360] Cohort 1: Phase 2a

[00361] Based on acceptable safety and reactogenicity (in the absence of any safety concerns or any events meeting a specific study pausing rule) as determined by DRC after the review of 14-day post-vaccination safety data for the initial 84 participants, the remaining 320 participants from Cohort 1 are randomized and dosed in Phase 2a of the study. These additional 320 participants were enrolled and randomized in parallel in a ratio of 2:2:2:1:1 to one of the five study

vaccination groups to receive a single IM injection of either ExPEC10V (1 or 3 doses), ExPEC4V or Prevnar 13 on Day 1 (Table 11).

In addition to performing the 14-day safety review for the initial 84 participants, the DRC also evaluates safety data of Cohort 1 over the course of the study and review any events that meet a specific study vaccination pausing rule or any other safety issue that may arise.

[00362] For Cohort 1, the primary analysis occurs when all participants have completed the Day 30 visit (Visit 4) or have discontinued earlier. The final analysis occurs when all participants have completed the Day 181 visit or have discontinued earlier. For participants progressing to the open-label long-term follow-up (LTFU) period (ExPEC10V selected dose group and Prevnar 13 group), yearly follow-up analyses include safety and immunogenicity data (multiplex ECL-based immunoassay and MOPA) collected up to the time of the visit at Year 1 (Day 366), Year 2 (Day 731) and Year 3 (Day 1096) after vaccination.

[00363] Cohort 2

[00364] In Cohort 2, the safety, reactogenicity, and immunogenicity of the selected dose of ExPEC10V (based on the primary analysis results of Cohort 1) is evaluated in participants aged ≥60 years in stable health with a history of UTI in the past 5 years. For Cohort 2, the study has a double-blind, placebo-controlled design, and a total of 600 participants are enrolled and randomized in parallel in a 2:1 ratio (400 participants in the ExPEC10V group and 200 in the placebo group).

[00365] All participants receive a single IM injection of either the selected dose of ExPEC10V or placebo on Day 1 per the assigned study vaccination groups (Table 11).

[00366] For Cohort 2, the primary analysis includes safety and immunogenicity data and occurs when all participants have completed the Day 30 visit (Visit 4) or have discontinued earlier. The final analysis occurs when all participants have completed the Day 181 visit or have discontinued earlier. For all participants, yearly follow-up analyses include safety and immunogenicity data (multiplex ECL-based immunoassay and MOPA) collected up to the time of the visit at Year 1 (Day 366), Year 2 (Day 731), and Year 3 (Day 1096) after vaccination.

[00367] A stool sample analysis is performed in a selected subset of participants to evaluate the effect of ExPEC10V on the prevalence of pathogens (eg, *Clostridium difficile*) and ExPEC10V serotypes in the intestinal flora using metagenomics.

[00368] NUMBER OF PARTICIPANTS

[00369] A total of 1004 participants is enrolled in the study; 404 participants in Cohort 1 and 600 participants in Cohort 2.

[00370] INTERVENTION GROUPS**[00371] Description of Interventions**

[00372] **ExPEC10V:** *E. coli* bioconjugate vaccine in phosphate buffered solution containing O-antigen PS of ExPEC serotypes O1A, O2, O4, O6A, O8, O15, O16, O18A, O25B and O75 separately bioconjugated to the EPA carrier protein. Single 0.5 mL IM (deltoid) injection of one of the three doses of ExPEC10V on Day 1.

[00373] **ExPEC4V:** *E. coli* bioconjugate vaccine in saline buffer solution containing O-antigen PS of ExPEC serotypes O1A, O2, O6A, O25B (4:4:4:8 µg PS/ExPEC serotypes) separately bioconjugated to the EPA carrier protein. Single 0.5 mL IM (deltoid) injection of ExPEC4V on Day 1.

[00374] **Prevnar 13:** Sterile suspension of saccharides of the capsular antigens of *Streptococcus pneumoniae* serotypes 1, 3, 4, 5, 6A, 6B, 7F, 9V, 14, 18C, 19A, 19F, and 23F, individually linked to non-toxic Diphtheria CRM197 protein. Single 0.5 mL IM (deltoid) injection on Day 1, supplied in a single-dose prefilled syringe.

[00375] **Placebo:** normal saline. Single 0.5 mL IM (deltoid) injection of placebo on Day 1.

[00376] The ExPEC study intervention materials are described in Table 9.

[00377] **Table 9.** BAC1001MV ExPEC Study Vaccines.

Study Arm	O1A (µg)	O2 (µg)	O4 (µg)	O6A (µg)	O8 (µg)	O15 (µg)	O16 (µg)	O18A (µg)	O25B (µg)	O75 (µg)	EPA (µg)	PS (Total) (µg)
Low dose ExPEC10V	4	4	4	4	4	4	4	4	8	4	160	44
Medium dose ExPEC10V	8	4	4	8	4	4	4	4	16	4	221	60
High dose ExPEC10V	8	8	8	8	8	8	8	8	16	8	320	88
ExPEC4V	4	4	-	4	-	-	-	-	8	-	72	20

EPA=a genetically detoxified form of exotoxin A derived from *Pseudomonas aeruginosa*; PS=polysaccharide

ExPEC4V consists of the O-antigen polysaccharides (PSs) of the ExPEC serotypes O1A, O2, O6A, and O25B separately bioconjugated to the EPA carrier protein.

ExPEC10V consists of the O-antigen polysaccharides (PSs) of the ExPEC serotypes O1A, O2, O4, O6A, O8, O15, O16, O18A, O25B and O75 separately bioconjugated to the EPA carrier protein.

Dose is based on PS only. The EPA (µg) are measured values.

[00378] ExPEC10V is composed of 10 monovalent drug substances (DSs). For this clinical study, 2 different concentrations (medium and high) of drug product (DP) are produced (Table 10). A third (low) concentration is obtained in the clinic by diluting the high concentration 1:1 with dilution buffer, which is the same as the formulation buffer. Each DP is formulated in Sodium/Potassium phosphate buffer at pH 7.0 (0.02% [w/w] Polysorbate 80, 5% [w/w] sorbitol, 10 mM methionine).

[00379] **Table 10:** Composition of ExPEC10V vaccine for phase 1/2a clinical study

Ingredient	Amount (µg/mL) ^a		
Active ^a	Low Concentration ^b	Medium Concentration	High Concentration
<i>O-antigen polysaccharide</i>			
EcoO1A	8	16	16
EcoO2	8	8	16
EcoO4	8	8	16
EcoO6A	8	16	16
EcoO8	8	8	16
EcoO15	8	8	16
EcoO16	8	8	16
EcoO18A	8	8	16
EcoO25B	16	32	32
EcoO75	8	8	16
<i>Carrier protein</i>			
EPA	320	441	640
Excipients			
KH ₂ PO ₄	6.19 mM		
Na ₂ HPO ₄	3.81 mM		
Sorbitol	5% (w/w)		
Methionine	10 mM		
Polysorbate 80	0.02% (w/w)		

EPA=genetically detoxified *P. aeruginosa* exotoxin A used as carrier protein

^a The active ingredient is a biologically synthesized conjugate composed of the PS antigen and a carrier protein (EPA); the dose is calculated on the PS moiety only.

^b The “low concentration” is obtained in the clinic by diluting the “high concentration” 1:1 with dilution buffer

[00380] SAFETY EVALUATIONS

[00381] Key safety assessments include solicited local and systemic AEs, unsolicited AEs, SAEs, physical examinations, vital sign measurements, and clinical laboratory tests.

[00382] IMMUNOGENICITY EVALUATIONS

[00383] Key immunogenicity assessments of collected sera include the assessment of ExPEC10V and ExPEC4V serotype-specific total IgG antibody levels elicited by the vaccine as measured by a multiplex ECL-based immunoassay, and ExPEC10V and ExPEC4V serotype-specific functional antibodies as measured by an opsonophagocytic killing assay (OPKA) in multiplex format (MOPA). Immunogenicity assessments of pneumococcal antibody titers elicited by Prevnar 13 are not performed.

[00384] The levels of serum antibodies induced by ExPEC10V are measured by a multiplex electrochemiluminescent (ECL)-based immunoassay. This assay combines high binding carbon electrodes in a multi-spot 96-well format microplate that is coated with different *E. coli* O-LPS antigens or the carrier protein EPA. The levels of antigen-specific antibodies present in serum samples are detected using a secondary antibody (anti-human IgG) labeled with SULFO-TAG.

The SULFO-TAG emits light in the presence of electrical stimulation at an intensity that increases proportionally to the amount of bound IgG antibodies. This assay was qualified according to International Conference on Harmonisation (ICH) recommendations.

[00385] The levels of functional antibodies induced by ExPEC10V are measured by a multiplex opsonophagocytic assay (MOPA). Briefly, heat-inactivated serum samples are serially diluted and incubated with different *E. coli* strains that are specifically resistant to different types of antibiotics. After that, human complement and phagocytic cells (HL60) are added to the reaction and, after a second incubation period, an aliquot of the reaction mix is transferred to different PVDF hydrophilic membrane filter plates containing media supplemented with specific antibiotic that selectively allow growth of a strain that is resistant to that particular antibiotic. After overnight grown, the colony forming units (CFUs) are counted to determine the number of surviving bacteria. This assay was qualified according to ICH recommendations.

[00386] For ExPEC10V serotype antibodies as measured by multiplex ECL-based immunoassay and MOPA, and EPA as measured by multiplex ECL-based immunoassay only, the following measures of immunogenicity are evaluated and tabulated by the study vaccination groups, for all immunogenicity time points:

- proportion of participants with a ≥ 2 -fold and ≥ 4 -fold increase in serum antibody titers from Day 1 (pre-vaccination)
- geometric mean titer (GMT)
- GMR: fold change from baseline, calculated from the post-baseline/baseline value.

For the LTFU period, descriptive summaries of immunogenicity are provided for each serotype.

[00387] Dose selection for later phases considers the totality of the evidence available at the time of the primary analysis of Cohort 1 (Day 30 results).

[00388] Table 11: Cohort 1: Vaccination Schedule

Study Vaccination Group	Vaccination on Day 1	Phase 1			Phase 2a			Total
		Step 1 Sentinel participants (Low dose)	Step 2 Additional participants (Low dose)	Step 3 Sentinel participants (Medium dose)	Step 4 Additional participants (Medium dose)	Step 5 Sentinel participants (High dose)	Step 6 Additional participants (High dose)	
G1	Low dose ExPEC10V*	2	18					80
	Medium dose ExPEC10V*			2	18			100
G2	High dose ExPEC10V*				2	18		80
G3	ExPEC10V*					2	18	100
G4	ExPEC4V**	1	3	1	3	1	3	40
G5	Prevnar 13***	1	3	1	3	1	3	40
Total		4	24	4	24	4	24	320
								404

* ExPEC10V consists of the O-antigen polysaccharides (PSs) of the ExPEC serotypes O1A, O2, O4, O6A, O8, O15, O16, O18A,

O25B and O75 separately bioconjugated to the carrier protein, a genetically detoxified form of exotoxin A (EPA) derived from *Pseudomonas aeruginosa*.

** ExPEC4V consists of the O-antigen polysaccharides (PSs) of the ExPEC serotypes O1A, O2, O6A, and O25B separately bioconjugated to the carrier protein, a genetically detoxified form of exotoxin A (EPA) derived from *Pseudomonas aeruginosa*.

*** Prevnar 13, Pneumococcal 13-valent conjugate vaccine (Diphtheria CRM197 protein) is a sterile suspension of saccharides of the capsular antigens of *Streptococcus pneumoniae* serotypes 1, 3, 4, 5, 6A, 6B, 7F, 9V, 14, 18C, 19A, 19F, and 23F, individually linked to non-toxic Diphtheria CRM197 protein.

[00389] Table 11: Cohort 2: Vaccination Schedule

Study Group	Vaccination on Day 1	Total
G6	ExPEC10V ^a	400
G7	Placebo	200
Total		600

^a ExPEC10V consists of the O-antigen polysaccharides (PSs) of the ExPEC serotypes O1A, O2, O4, O6A, O8, O15, O16, O18A, O25B, and O75 separately bioconjugated to the carrier protein, a genetically detoxified form of exotoxin A (EPA) derived from *Pseudomonas aeruginosa*.

[00390] The randomization ratio for the participants enrolled in Cohort 2 of the study is 2:1 (ExPEC10V:Placebo). The ExPEC10V dose used in Cohort 2 is based on the primary analysis (Day 30) results of Cohort 1.

[00391] STATUS

[00392] Enrollment and vaccination of Cohort 1 of the study described above was completed. The study is ongoing in a blinded manner. Based on ongoing review of the safety data, no major safety issues were identified, and the ExPEC10V vaccine has an acceptable safety profile.

[00393] The analysis of the immunogenicity of the Cohort 1 clinical samples is ongoing in a blinded fashion. The ECL data were 100% Acceptance Quality Limits (AQL) checked and uploaded for data management. Analysis of the MOPA samples is ongoing. Data unblinding and statistical analysis is performed by using a clinical research organization (CRO).

[00394] The Cohort 2 vaccinations are started once the ExPEC10V dose for that Cohort has been identified based on the finalized primary analysis of the Day 30 results from Cohort 1.

[00395]

[00396] It will be appreciated by those skilled in the art that changes could be made to the embodiments described above without departing from the broad inventive concept thereof. It is understood, therefore, that this invention is not limited to the particular embodiments disclosed, but it is intended to cover modifications within the spirit and scope of the present invention as defined by the present description.

PAGE INTENTIONALLY LEFT BLANK

SEQUENCES

SEQ ID NO: 1 (Glycosylation consensus sequence)

Asn-X-Ser(Thr), wherein X can be any amino acid except Pro

SEQ ID NO: 2 (Optimized glycosylation consensus sequence)

Asp(Glu)-X-Asn-Z-Ser(Thr), wherein X and Z are independently selected from any amino acid except Pro

SEQ ID NO: 3 (EPA carrier protein comprising 4 glycosylation consensus sequences (EPA-4))

G SGGGDQNATG SGGGKLAEEA FDLWNECAKA CVLDLKDGRV SSRMSVDPAI ADTNGQGVHL YSMVLEGGND
 ALKLAIDNAL SITSDEGLTIR LEGGVEPNKP VRYSYTRQAR GSWSLNWLVP IGHEKPSNIK VFIHELNAGN
 QLSHMSPIYT IEMGDDELLAK LARDATFFVR AHESNEMQPT LAISHAGVSV VMAQAAQPRRE KRWSEWASGK
 VLCLLDPLDG VYNYLAQQRC NLDDTWEGKI YRVLAGNPAK HDLDIKDNNN STPTVISHRL HFPEGGSLAA
 LTAHQACHLP LEAFTRHRQP RGWEQLEQCG YPVQRLVALY LAARLSWNQV DQVIRNALAS PGSGGDLGEA
 IREQPEQARL ALTLAAAESF RFVRQGTGND EAGAASADVV SLTCPVAKDQ NRTKGECAFP ADSGDALLER
 NYPTGAEFLG DGGDVSFSTR GTQNWTVERL LQAHRQLEER GYVFVGYHGT FLEAAQSIVF GGVRARSQDL
 DAIWRGFYIA GDPALAYGYA QDQEPMARGR IRNGALLRVY VPRWSLPGFY RTGLTLAAPE AAGEVERLIG
 HPLPLRLDAI TGPEEEGGRV TILGWPLAER TVVIPSIAPT DPRNVGGDLD PSSIPDKEQA ISALPDYASQ
 PGKPPREDLK LGSGGGDQNA **T**

SEQ ID NO: 4 (O4 GtrS amino acid sequence)

MNNLIMNNWCKLSIFIIFILLWLRRPDILTNAQFWAEDSVFWYKDAYENGFLSSLTPRNGYFQT
 VGLTALLNPDYAPFVNFGIMIRSVIIWFLFTERFNFLTTRIFLISYFLCMPGLDEVHANITNAHWYL
 SLYVSMILIARNPSSKSWRFHDIFFILLSGLSPGPFIIFILAASCFKFINNCKDHISVRSFINFYLRQPYAL
 MIVCALIQGTSIILTFNGTRSSAPLGFSFDVISSIISNSNIFLFTFVPWDIAKAGWDNLLSYFLSVSILSC
 AAFVFVKGTWRMKVFATLPLLIIIFSMAKPQLTDSAPQLPTLINGQGSRYFVNIIHIAIFSLLCVYLLECVR
 GKVATLFSKIYLTILLFVMGCLNFVITPLPNMNWREGATLINNAKTGDVISIQVLPPGLTLELRKK

SEQ ID NO: 5 (Example O4 *gtrS* nucleic acid sequence)

ATGAATAATTAAATTATGAATAACTGGTGAAATTATCTATATTATTATTGCATTATTGCTATGGCT
 TAGAAGGCCGGATATACTCACAAACGCACAATTGGGCAGAAGATCCGTTCTGGTATAAGGACGCCT
 ATGAGAACGGATTCTTAAGTTCACTAACAAACGCCAGGAATGGGTATTCCAGACTGTTCTACATTATA
 GTTGGTCTGACTGCTTATTAAATCCAGATTATGCACCTTTGTTCTAATTGGCATAATGATTG
 CTCAGTAATTATATGGTTTATTACAGAAAGATTCAACTCCTCACATTGACTACTAGGATTCTTAT
 CTATTATTCTATGCATGCCGGATTGGATGAAGTTCATGCAAATATAACAAATGCACATTGGTATTG
 TCATTATATGTATCAATGATCCTGATAGCTCGCAATCCAAGTTCAAATCATGGAGGTTCATGATATATT
 CTTTATCTGCTATCCGGGCTCAGTGGCCCATTTATAATTTCATTAGCAGCTTCATGCTTAAATT
 TAAATAATTGTAAGATCATATTAGTGTAAAGATCTTCAAAATTCTACTTGCCTCAGCCATACGCATTA
 ATGATTGTTGCGTTAATTCAAGGAACCTCTATAATTCTAACCTTCAATGGCACACGTTCTCAGCACC

GCTAGGATTCAAGTTGATGTGATTTCGTCTATTATATCATCGAATATTTTATTACATTGTCCCCAT
 GGGATATTGCAAAGGCTGGGTGGGATAATTTACTGTTATCTTATTTTGCTGTTCGATTTGTCGTGT
 GCGGCCCTTGTGTTAAAGGTACGTGGCGAATGAAAGTATTTGCAACTTACCATGCTAATTATAAT
 ATTTCAATGGAAAACCACAATTGACAGACTCGGCACCTCAATTGCCAACACTTATTAATGGGCAAGGTT
 CAAGATACTTCGTAAAATATACATATTGCGATATTCTCTTGCTATGTGTTACTTACTTGAGTGCCTCAGG
 GGGAAAGTGGCAACTTATTTCCAAAATATACTTAACAATTTGCTATTCGATGGGATGTTGAATT
 TGTTATCACCCCACTCCCAAACATGAACGGAGGGAGGTGCTACTTGATTAAATAATGCAAAAACGGTG
 ATGTCATTCGATTCAAGTGTACCACTGGCCTAACACTTGAACTAAGGAAAAATAA

SEQ ID NO: 6 (Example PgIB sequence ('wild-type'))

MLKKEYLKNPYLVLFAMIILAYVFSVFCRFYWWWWASEFNEYFFNNQLMIISNDGYAFAEGRDMLIAGFHQ
 PNDSLSSYGGSSLSALTYWLYKITPFSFESIILYMSTFLSSLVVIPTILLANEYKRPLMGFVAALLASIANSY
 YNRTMSGYYDTDMLVIVLPMFILFFMVRMILKKDFFSLIALPLFIGIYLWWYPSSYTLNVALIGLFLIYTL
 IFHRKEKIFYIAVILSSLTLSNIAWFYQSIIIVILFALFALEQKRLNFMIIIGILGSATLIFLILSGGVDP
 IYQLKFYIFRSDESANLTQGFMYFNVNQTIQEVENVDLSEFMRRISGSEIVFLSLFGFVWLLRKHKSIM
 ALPIVLVGLFLALKGGLRTIYSVPVMALGFGFLLSEFKAIMVKKYSQTSNVCIVFATILTLPVFIHIYN
 YKAPTVFSQNEASLLNQLKNIANREDYVVTWWDYGPVRYYSDVKTIVDGGKHKGKDNNFFPSFALKSDEQA
 AANMARLSVEYTEKSFYAPQNDILKTDILQAMMKDYNQSNVDLFLASLSKPDKIDTPKTRDIYLYMPARM
 SLIFSTVASFSFINLDTGVLDKPFNSTAYPLDVKNGEIYLSNGVVLSDLFRSKIGDNVVSVNSIVEINS
 IKQGEYKITPIDDKAQFYIFYLKDSAIPYAQFILMDKTMFNSAYVQMFGLGNYDKNLFDLVIINSRDAKVFK
 LKI

SEQ ID NO: 7 (example gtrA amino acid sequence; *E. coli* W3110 yfdG, GenBank: BAA16209.1)

MLKLFakytsigvlntlihwvvfgvciyvahtnqalanfagfvavsfssfanakftkastttmrymlv
 gfmgtlsatvgwaadrcalppmitlvtfsaislvcgvfyksfivfrdak

SEQ ID NO: 8 (example gtrB amino acid sequence – *E. coli* W3110 yfdH, GenBank: BAA16210.1)

MKISLVPVFNEEEAIPIFYKTVREFEELKSYEVEIVFINDGSKDATESIINALAVSDPLVVPLSFTRNFG
 KEPALFAGLDHATGDAIIPIDVDLQDPIEVIPHIEKWQAGADMVLAKRSDRSTDGRLKRTAEWFYKLHN
 KISNPKIEENVGDFRILMSRDVVENIKLMPERNLFMKGILSWVGKTDIVEYVRAERIAGDTKFNGWKLWNL
 ALEGITSFSTFPLRIWTYIGLvvvasvafiygawmildtiifgnavrgypslvsilflggiqmiwigvlge
 yigrtiyietkkrrpkwyikrvkk

SEQ ID NO: 9 (example O4 rfb locus nucleotide sequence – O4-EPA production strain BVEC-L-00684f)

ATGACGAATTAAAAGCAGTTATTCTGTAGCGGGCTCGGGATGCATATGTTGCCTGCCACTAAGGCGAT
 ACCCAAAGAGATGCTACCAATCGTGACAAGCCAATGATCAGTACATTGTTGACGAGATTGTCGCTGCAG
 GGATCAAAGAAATCCTCTGGTAACTCACCGTCCAAGAACGCGTCGAAAACCACTTCGACACCTCTTAT
 GAGTTAGAATCACTCCTGAGCAGCGCTGAAGCGTCAACTGCTGGCGGAAGTACAGTCCATCTGTCGCC
 GGGCGTGACCATTATGAACGTGCGTCAGGGCGAACCTTAGGTTAGGCCACTCCATTTGTGTGCGCAG
 CTGCCATTGGTGACAACCCATTGTCGTGGTACTGCCAGACGTTGATCGACGATGCCAGCGCCGACCCG
 CTACGTTACAACCTGCTGCCATGATTGACAGTTCAACGAAACGGCCGCAAGCAGGTGCTGGCAAAACG
 TATGCCGGGTGACCTCTGAAATCTCCGTATCCAGACTAAAGAGCCGCTGGACCGTGAGGGTAAAGTCA
 GCCGCATTGTTGAATTATCGAAAACGGATCAGCCGAGACGCTGGACTCAGACATCATGGCGTAGGT
 CGCTATGTGCTTCTGCCGATATTGGCCGGAACTGGAACGTACTCAGCCTGGCATGGGACGTATTCA
 GCTGACTGATGCTATTGCGAGCTGGCAAAACAAATCCGTTGATGCAATGCTGATGACCGCGACAGTT
 ACGACTGCGCAAAACATGGGCTATATGCAGGCCTTGTGAAGTATGGCCTACGCAACCTGAAAGAAGGG
 GCGAAGTCCGAAAGGTATTGAGAAGCTGTTAAGCGAATAATGAAACCTGACCGGATGTAACGGTTGAT
 AAGAAAATTATAACGGCAGTGAACATTGCGAGCAAAAGTAATTGTCGAATCTCCTGCCGTTTTA

TATAAACCATCAGAATAACAACGAGTTAGCAGTAGGGTTTATTCAAAGTTCCAGGATTTCTTGTT
CCAGAGCGGATTGTAAGACAATTAGCGTTGAATTTCGGGTTAGCGCAGTGGGTAAACGCTCGTCAC
ATCATAGGCATGCATGCAGTGCCTGGTAGCTGAAAGCAGGGCGTAGCGTCATTAAACCTCTATT
AATCAAACGTAGAGGCCGCTTATTACAGCATGCTGAAGTAATATGAAATAAAATTAAAGTAAAATACTT
GTTACTGGTGGCGCAGGATTATTGGTTAGCTGAGCTGACATTACAGGCTGATGTTCTGATTCTGAACGCT
ATGTTTGAAACATGCAGGATATTGCGATGCACCTGCAATGGCACGGATTGGCTCAGCATCAGCCGGAT
GCAGTGATGCACCTGGCTGAAAGCCATGTTGACCCTCAATTACAGGCCCTGCGGCATTATTGAAAC
CAATATTGTTGGTACTTATGTCCTTGGAAAGCCGCTCGAATTACTGGTCTGCTCTGATAGCGACAAGA
AAAATAGCTCCGTTTCATCATATTCTACTGACGAAGTATATGGTATTGCTCATCCTGACGAGGTA
AATAATACAGAAGAATTACCCATTAACTGAGACAACAGCTACCGGCCAAGCAGCCATTCCGCATC
CAAAGCATCCAGCGATCATTAGTCGCGCTGGAAACGTACCTATGGTTACCGACCATTGTGACTAATT
GCTCTAACAAATTATGGTCCTTATCATTCCGGAAAATTGATTCCATTGGTATTCTCAATGCTCTGGAA
GGTAAAGCATTACCTATTATGGTAAGGGGATCAAATTGCGACTGGCTGTATGTTGAAGATCATGCGCG
TGCCTTATATACCGCTAACCGAAGGTAAGCGGGTAAACCTTATAACATTGGTGGGCACAACGAAAAGA
AAAACATAGATGTAGTGCCTACTATTGTGATTGCTGGATGAGATTGACCGAAAGAGAAAATCTTATCGT
GAGCAAATCACTTATGTTGCCATGTCGGGACACGATGCCGTTATGCGATTGATGCTGAGAATTGG
TCGCGAATTGGGATGAAACCACAGGAAACGTTGAGAGCGGGATTCCAAGACAGTGGATGGTATCTGT
CCAATACAAAATGGGTTGATAATGTGAAAAGTGGTGCCTATCAATCGTGGATTGAAGAGAAACTATGAGGGC
CGCCAGTAATGAATATCCTCCTTTGGAAAACAGGGCAGGTAGGGTGGGAACTACAGCGTCTGGCA
CCTCTGGGTAACCTGATTGCTCTTGTATGTTCACTGATTATTGTGGCATTTCAGTAACCCGAAGG
TGTGGCTGAAACCGTCAAAAAATTGCGCCAGATGTTATGTTAATGCTGCTCATACCGCGTAGATA
AGGCTGAGTCAGAACCGAGATTGACAAATTACTCAATGCGACCAGCGTTGAAGCAATTGCAAAAGCGGCT
AATGAAGTTGGGCTTGGGTAATTCAATTACTCAACTGACTACGTCTCCCTGAAATGGCGACATGCCATG
GCTCGAGACTGATGTAACCGCTCCGCTCAATGTTATGCCAAAACCAATTGGCTGGAGAAAAGAGCATTAC
AAGAACATTGCGCAAAGCATCTTATTCGTAACCGAGCTGGTATATGCAAGGTAAGGAAATAACTTGC
AAAACAATGTTACGCTGGCAAAAGAGCGCAGAACTGGCTGTATAACGATCAGTTGGCGACCAAC
AGGTGCTGAATTGCTGGCTGATTGACCGCTATGCCATTGCGTGGCATTAAAAACAGAAGTTGCTG
GCTTGTACCATCTGGTAGCAAATGCCACAACACCTGGCACGATTACGCCGCTAGTATCGAAGAAGCC
CGTAAAGCAGGGATTGACCTGACTAACAAACTCAACGCCGTTACCAACACGGCTTACCTCAGC
CCGCCGTCCTCATAAATTGCGCTCAATACCGAAAAGTTCACTGAGAACTTGGCCTGCTGCCTGACT
GGCAGGTGGCGTAGAACGTATGCTAACGAAATTACGACTACGGCAATTAAACAAATTGGCATCT
CGCTCATGATGCCAGAGCGGGATGAATTAAAGGAATGGTAAAGACGCTAAAGGTATTCTGG
CTGGTGGTCCGGCACTCGTCTTATCCTGTGACGATGGCAGTGAGTAACAAACTGCTGCCGATTATGAT
AAGCCGATGATTATTATCCGCTTCAACGCTATGTTAGCGGGTATTGCGATATTCTTATTACAGTAC
GCCACAGGATACACCGCGTTCCAACAATTGTTGGGGACGGGAGTCAGTGGGCTTAATCTACAGTATA
AAAGTACAACCGAGTCCGGATGGCCTGGCGCAAGCGTTATTATGGTGAAGACTTATTGGTGGTGTGATGAT
TGTGCACTCGTACTTGGCGATAATATCTTCTATGGACACGACTTGCAGAAATTATGGTGTGCTGTTAA
CAAAGAAATCGGTGCAACGGTATTGCTTATCAGTCATGATCCTGAACGTTATGGTGTGAGTTG
ATAATAACGGTACTGCAATTAGCCTGGAAGAAAAACCGCTGGAACCAAAAGTAACTATGCGGTTACTGG
CTTTATTCTATGACAATGATGTTGAGAAATGGCAAAAACCTTAAGCCTCTGCCGTGGCAACTGG
AATTACCGATAATTAAACCGTATTATGGAGCAGGGACGTTGCTGCTATGATGGGCGTGGTATG
CCTGGTTGGATACTGGTACACATCAAAGTCTTATTGAAGCAAGTAACCTCATTGCCACCATGAAGAGCGT
CAGGGATTAAAGGTATCTGCCGGAGAGATGCTTACCGTAAAGGGTTATTGATGCTGAGCAGGTGAA
AGTATTAGCCGAACCGCTGAAGAAAAATGATTATGGTCACTGCTAAAGGTTATTGATGCTGTTAA
AAAATGAACGTAATTAAACGAAATTCTGATGCTGATTGGAAACCAAAAGTTTGGTGTGAAACG
TGGCTTCTTTGAGAGTTAACGAAAGTATTGAAGAAGCTGAGGACGGAGGTTGAATTGTT
AGGATAACCATTCAAGTCTAAATAATGTATTGCGTGGGATGCATTATCAAACACAAAATCTCAAGGA
AAACTGGTCCGGTAATTCTGGTTCAGTATATGATGTTGCCAGATTAAAGAGAAAATCAAAGACATT
TGGCAAATGGTGGGTGAGAATTATCTGGGATAATAAAAGACAATTGCGATCCCAGAGGTTTGGCC
ATGGTTTTATGTTGGAGGAGAATACCGAATTGTTATAATGTACCGATACTTATAACCTGCTCAT
GAACACACATTGCTATGGAATGATCCAACATCAATATAAGTGGCCAATCATACAAACTGCAAGCCAAT
TATTCTGAAAAAGATGCTAATGGACATCTTTCACATAAAACCTATTCTGAAATGCAATATTGAG
TTAATTAGAACAGTTCTATAATTGCTGGTTGCTGCGACATTAGTGCAGTCCCTGCTTGG
GGATTCTGCCAGGCTGCTGGACGGAGAATTGGACTTTCACACTAGCATTGCTTGTAGGATAT

GCAAGTATTCGACGCCGGATTAGTCGAGCTGAATCAGAGAAATCGCTTTATCGAGAAAGTGAAAA
 AGAGCAAATAAAATTATTCGACAGCAAGTGTAAATCGTACTATTCTAGGGGTGGTGCAGCTTGTAC
 TTTATTTAGTAGTAATAAAGTTGTGAGTTATGAAATGTTAGTCGTTATATTGAAACAGCAGTGCCT
 GCATTCTCTGTATTCATTATAAACCTGTATCTGATTAACCAGATTGGCTGGTATCTGGCTATTACCA
 GCTAGAAAATTGCAAATAAAATGTTAGAGAATGATTCTAGCACAAGCTGGCTATTACCAAGTGA
 TATTTGTTATTACAATCCCTCGTTGCTTATGCTATGTATGGTTGGTGGCGTGTGATTCTATT
 TTGATTAGCGAATAATTGTCGAGATATTCTTAAAGTAAACTTACTTTAATGTGGCAACTTGCA
 TCGCTTATCTCTTGGGATGGATAACAGTTAGTAATATCATAAGCCAATCATGGCATATTGCA
 GCTTATCATCTCATATTATGGGGCTCGAGAATTGCATTACAGGCCCTCAGAGGGTGTATCA
 AGGTTAATTAAATAATCCCATACTGCTTGGCAAGAGCTCTATTCTAAATTGGCATATAGCAATA
 TGAAACGAAAAAAATTACAACACAGAGCTACGCAATTATAAGCATTGTATGTACCCATAGTTG
 GTGTCATTTGCCTCATTCATAATGACAACATGGATGGGACCTGATTATGCCTAGAAGCAGCA
 ATGAAAATACTCTTGCTGGTTTTCTTAACTCTTAGCGCAAATACTTATGCATACTTGCA
 CGGAAAGTCAAAATTACCGCATTGTCATCTAGAAACTTGCGCCACTATTATTATTGTATT
 TCACAAATGCATTGCGATAATTGGCACGGCAATCGCTGGTCACTTAGAACATTGTTGATT
 CTACTTTCGATATCGAGAAGAAAATGATTGCGTTGATATTGCGCTTGCAACCTACA
 ATTGGCAACAGATTGAATCTATCCAGAAACAAACTTATAGAAATTGGCGTCTTATAA
 AGTGTGATGATACTGTTGATATTAAAGGATAATGATGTCTAACGACAGTGTATCTATT
 CTCGAGTGAAGGAGGGTTATTCAAGAACCTTAATTATGCTCTTCACAAACTACATCT
 GAAATTGTTGATTAGGAAATTAGACAGAAATAATTGTTGATTG
 CAGAAAGTTTATCGAACAAATAATTAAATCCACAAGATAATCTGAAAATAATAAT
 CTTCTGGCGTCAACGGTATATGGCTGACTTGCATCATGAATAAGAAACTTGTGATATT
 GCATTGCGCTATACCTACATA
 TGCACATATGCATGATCAATGGTGGCATTATTAGCGAACATATGGTAACATT
 CGTCTGTTCGTTAGGCAACATTCTACAAATGTTGTTGGTAGAAATAACAGC
 CATTCAAAATT
 AATTCCATACAAAAAAACTAAAAAGGATTAAATTGCTAGTGGATAGAAC
 ACTGTTGCTTAAATTAAATCAAA
 TAACGATTCTATCCAGGAATAAAATGGAAAATAAAATTGATTACT
 TACCTTATCTTTAAAGGAAACAAGAAAAGTTTCACTTGTGATTAAATT
 TGATATATTATTATTGCACTGTTATGATCTGACGTTTAACACACAGGG
 GACAGGCATT
 TATGTTGATCTGCGTTAGTATTCTTTGGCTTAACCTATCCATCAGGAGGG
 ACTGGATAGGTTA
 TTTCTCCATTATGACTGCATGGTAATGAGCAGTGTAAATGGTT
 ATAATGTTGAACCTGGATATG
 AATTAAATTGTTCCATTGGATATTGGGATTTCAGACAATT
 ATTATTATAGCCGTGAAAGTAA
 ATTCTAATATTAAATTGCAAAGCATTGAAAACGGAAAGTTTGTATT
 GTGCGATAATGTGCATGTT
 CCTTGGAGTGTATTGAGGCATTAGACAGGCTCTGGCCTATCT
 ATAGTTATTGGGATTCT
 CTCTTTTTGGTAGAAAAGGAAATTATAACATTAGTATT
 TGCGTCAACTTCCATATAACTGCT
 TTGATTGTTCTTAATGACTCCTCTATT
 CAAGAAATTAAAGCAAGATAATAAGTTAGCCTATT
 AATTTCAGTAGCTCTTTCGCTTTCTGAAACCATAATTAAAGTC
 ACTCCTGCAATTGCGAG
 GATCCATTGCCAGTGAAAATTAAAGTTTACTTAGCAACCGAGCA
 ATACAGGCCACAGTTATCTATTGG
 AGTGCACATTCTGACATTATAACTTATT
 TTCTGATATGTGTAAGTTAACGAATAAGAAATAT
 GCTCGCTAATTATAATGCTGCAAATGAGATATTGCTTATT
 GGTTGCTGTCTTATATTCTCGGTATT
 TTATCGGGAAATGATGCCAGTTATGACTCGCATTGGTTGGTATGG
 TTTCCATTGTTATAGTACTCTT
 TATAATTGCTTTATTGGCTACAAATT
 TGCGACCATTAAACATATGATT
 TAGCTATTATAATATAA
 TGCA
 CACCAGGACTTTGCTGAATAGGTTGATGCATTAGATGATGC
 ATTAAGACAATCAGCGAAGAGA
 AAATGTTGATTGGAAAGATAGGATATGGTTCTTATGTA
 GATGTTGAAATT
 ATTGTTAATAAGCATTGCTCTGCTCTGCTCTGCTTGA
 AACATTGTTGAAATT
 AGATAACCACTAGGA
 ACTGTATGTTGATCTG
 CAAAAATT
 ATTGTAAGTGC
 GACGGCGCTGG
 TTCCGGAGGTGC
 ATTAACATA
 TAAAGCAATT
 ATTAAA
 ACATGC
 ATCACAA
 ATTCAA
 ATGACT
 ATT
 TGTTGATCTGCGGGATTGGAGTTGCGCGGTCTGT
 GATAACATCAT
 TTACATAGAAA
 ACACAC
 CAAAAGGA
 TGGTGAAAAGA
 ATATGGGATTGGT
 CGGGTGT
 CGGAAGTT
 ATCTCG
 GAACATAAG
 ATTACGTTAA
 GAAAGTAATT
 TCTCTAC
 AAAATT
 CCAGTT
 GATGTT
 GAATGTT
 CCTTAC
 GAACAG
 ATT
 ATTACT
 TGC
 ACCAG
 CCAA
 TTC
 CCTT
 TAGTAAAG
 ATTG
 ATT
 TTT
 TAT
 ATT
 TAA
 ATT
 ATG
 GAATG
 CCA
 ATAC
 AAC
 CAT
 CGT
 ACTG
 GCA
 AAC
 GAATT
 GGAT
 GTAT
 CTG
 CAAA
 AAC
 ACT
 CCT
 TAT
 ATCC
 AGCG
 ACA

CCACTTACCTATAAAATCATTGGTCATTCTGAAGGC GTGGTTTTAAAGAAAAAGTATTTATAGA
 TGATCTGAAATCCAAGTGACTTTGAAAAGAATAGGTACAAAATTTGATAAGTTGTCAATTAAATA
 ACTTAAGCAAAACGTTGATTATCTGGCGTCTTCATACTGAAC TTGCAAAAAAATATGGCGCA
 TCTTAATCGTTTCCTAGCTATCGAATCATATGGGTACCACTCATCGAAGCTGCTAGTTAGGAA
 AAAATCATTAGTAGTGTACTTCCTATGCCCGGGATGTTAAAGGATTATAGCGCGTAGATTGTAA
 TTTACAATAATGAAGATGGCTGGCTAAGCGTTGTTAATGTTAAATGGCAATT CGAAGCTCAATT
 AGGCCTATGAAAAGATAGTCGTTCATCTGCCACAGTCTCTATTGAAATAAGGTGTATTATG
 TTTAATGGTAAAATATTGTTAATTACTGGTGTACGGGCTT CGTAATGCTGTTCAAGACGTTCT
 TGACACTGATATCAAAGAAATACGTATTTCCCGGGATGAAAAAAACAAGATGACATGAGGAAAAA
 ATAATAATCCGAAACTTAAGTCTATAGGTGATGTTCGCAGTATT CGAGTATCCTCAATGCTCTCGA
 GGTGTTGATT TATTTATGCTGCACTGAGCTGAAGCAAGTACCTCCTCGCAATTCCACCCAAATGGAAGC
 TGAAAAACGAATGTTAGGTACGGAAAACGTA CTGGAAGCGGCAATAGCTAATGGAGTTAGGCAGATTG
 TATGTTGAGTACAGATAAAGCTGTATATCCTATCAATGCAATGGGTATT CCAAAGCGATGATGGAAAAA
 GTAATGGTAGCAAATCGCGCAATGTTGACTGCTCTAAACGGTTATTGCGGTACACGTATGGCAATGT
 AATGGCATCTCGTGGTTCA GTTATCCATTATTGCGATCTGATTAATCAGGTAGACCAATGACGATAA
 CAGACCCTAATATGACTCGTTCATGACTCTCGAAGACGCTGTGATTGGTTCTTACGCATTGAA
 CATGGCAATAATGGTGTATTTGTCAAAAGGCACCTGCGGCTACCATCGAACACGTTGGCTATTGCACT
 CAAAGAATTACTTAATGTAACCAACACCTGTAATATACTGGCACCCGACACGGGGAAAACGTACG
 AAGCGTTATTGAGCCGAGAGGAATGATTGCA CGCGAGGATATGGGTGATTATTATCGTGTCCACCA
 CTCCCGGATTTGAACTATGAAAATATGTTGAAACATGGT GACCGTGTATCTGGAAAGTGGAAAGATTATAA
 CTCTCATAAACTGATAGGTTAGATGTTGAGGGAATGAAAAAAATTACTGCTAAACTTCCTTTATCCGG
 CACTCGGTCTGGTGAAGATTATGAGTTGGATCATAATATGAAAATT TAGTTACTGGCGCTGCAGGGTT
 TATCGGTGAAATTGGTATTCCGGCTTAAGGAAGCTGGATATAACGAACTCATTACGATAGATCGTA
 ACTCTCTGGCGGATTTAGAGCAGGGACTTAAGCAGGCA GATT ATTTCACCTGCTGGGTTAAATCGT
 CCCGTGAAGGAGTGTGAATTGAAAGAGGGAAATAGTAATCTAACTCAACAGATTGTTGATATCCTGAA
 AAACAATAAAATACTCCATCATGCTGAGTTCTCATCCAGGCTGAATGTGATAACGCTTATGGAAAGA
 GTAAAGCAGCTCGGGAAAAAATCATCAGCAGTATGGGAAACGACAAACGCTAAATATTATTCATGC
 TTGCCAATGTATT CGTAAGTGGTGTGACCAAATTATAACTCCTTATAGCAACTTCTGCCATCGC
 TGCAATGATGAAGCTATTACAATTATGATCCTCAGCAGTTGTAATCTGGTGTATAGATGACTTT
 GTTCTGACATATTAAAGCTATTAGAAGGAGCGAACGAAACTGGTACAGGACATTGGTCCAATTATTCT
 GTTACTGTTGGTGAAGTGGCACAATTAAATTACCGGTTAAAGAAAGTCGCCAACATTAATCACC
 GAAGA TGTAGGTAATGGATTACACGTGCATTGACTCAACATGGTAAGTTACCTGTCCTGAACAGTT
 GCGTATACCGTTCTTATAGTGATGACAGAGGGTATTCTGTGAAGTATTGAAAACGAAAACGCGGCCAG
 TTTCGTTCTTACTGCGCATCCAGGAATTACTCGGGTGGTCATTATCATCATTCCAAAATGAGAAATT
 TATTGTCATCGGAGGAAGTGCTTCAAAATTGAAAATTGTCACGAGTGAACGATATGAACCTTAATG
 TTCCCTGATGATTAAAATTGTTGAAACAGTTCCGGGATGGACGCATAACATTACTAATAATGGCTCG
 GATGAGCTAGTTGTTATGCTTGGGCAAATGAAAATT TAATCGTCTGAACCA GATACTATAGCGAGAGT
 TTTATCGTAAAAAATTGAAAGTCATGTCGGTTGTTGGGACTCGTCAGAAATTATCGACTCTCGCGTGT
 CCTTGCAAAATTAGATGAATATTGTCACCCATTGTCATACCGGGCAAAACTACGATTATGAAC
 ATGAAGTTTTCAAAAGATTGGGTGTCGCAAAACCTGATTATTCTTAATGCCGAGGTAAAATGCA
 GCAGAGACTATTGACAAGTTATCATTAAAGTGATGAGGCTCTGAAACAGGAAAACCAGAACG
 CCATGTTAGTGGCGATACTAACTCCTGTATTCAACGCAATACCA
 CAGCAAAGCGTGAAGAATTCCGATCTCCATA
 TGGAGGCTGGGATCGTGTGACCAACCGGTACCGGAAGAAACTAACAGAAAATAGTGATCATACC
 GCTGATATCAATATGACATATAGTGATATCGCGTGAATATCTCTGGCTGAAGGTGTACCGCC
 GATAGAATTATAAAACCGGTAGCCAAATGTTGAAGTACTCACTCATTATATGCCGACATTGATGG
 TCCGATGTTCTCGCCTGAAATTAAACACCTGGGAAATTCTTGTTGTTGAGTGC
 CCACAGAGAAGAAAATGTTGAT ACCCTAAACACTTGTGAAACTGGCAATATACTTAATACCG
 TGGCTGCTGAAATGATGTTGACGTTACGTTAAACGGGTGTTGAGAAGCAATACTA
 ATGAAACTTGCAATTATGATGATTATGAGTATTGCCCCATAGC
 ACACGCGTTGGGCTAAAATGATGTCCCCTGAGT
 TGTTCTACTCATCCTCGCACTCGTAACCGCATCAACGAAAACGGTATT
 CAATTCCATAAAAATATCTTGC
 TTCTTAAGCCATTAGGATTTCACGATTACAACCATCTGCAAAA
 ATGCACTGCGTGTGTTATCGGATAGT
 GGGACTATTACAGAACAGTCCTCCATTATGAACTTCC
 CTGCACTCAATATACGAGAACGCGCACG
 GGAAGGCTCGAAGAACGGGAGTAATGATGGTGGCT
 CTGAATCTGATCGCGTTTACAGGATTAG
 TAAAGTGTGACTATAGCATGCCAATGTT
 TCAGATAAAAGTCTCGGTATTATCCATT
 CATATACTGACTACGTTAAACGGGTGTTGAGAAGCA
 ATACTAATGAAACTTGCAATT
 ATGACGAGGCACTGTAATT
 ACGCGCTGACATCTCATTACAAGCAA

TTTATTCTATTAGTATGATTGATGGTATAAAGGTTGGCGTTCAAAAGTGGACCTTAAAGGATGTAGGT
 AAGGCTAACGTGCCATAATGAAACTCTTTATCTTCGCGCATGGCGCATTAAAGCACCTCATTCA
 ACATGATACATTGATGGTATCGTTATTATCCCCCTCATTTTGCGACTTGGTAAAAAAATAA
 AACACGATGCCAGTGCCAAGCTATCTGATCCAAGGGATATGTTCCACAGTGGTCATTGATGCAGGT
 ATGTTGAAAGCCGGTCACCAATTGAAAAATATTAGGTATTGAAAAAAAGTCATATCAGCAGGCTGG
 CCGGATAGGGTAATGTCGATAAGAATCTGAGATATTGCCAGACCAATAAGGTTACCGTGTGAAG
 TTTACGTAATTGGCCTCAATGACTCCTGTGCTGCCAGCGATGATTATCATCACTCGTCAAAATAC
 GATCTAAAAGATAAAAGTCATTTTCTATGGCGTAATATTGGCATGCTCAGGATATGGCAAACCTAAT
 GCGCCTGCGCTAATATGATGCGTATCATGATGCTATTCCCTGTTATAGGGCAGGGTATGAAGTTG
 AGCTGATAAAATCTTGCAGAATGGAATTAACTAATTCACTCATCTACCTCAGTGAACCAGGAA
 GAGTTAAATTAAATTCTGAAGTTGATGTCGGCCTGTTCTCCCTTCATCTGCCATTTCACATAA
 TTTCCCGGAAAATTACTAGGGTATATGGTTCAATCAATCCGATCCTGGGAGTGTGAATGGCGGAATG
 ATTTAATGGATGTAATTAAATAAGCACAGAGCCGGTTCAATTGATGTTAATGGTGAAGATGATAAACTGTT
 GAATCTGCACAATTGCTCTTAGTGAATTGACTGTTAAGAAAACAGCTAGGTAGTCAGAACGCTAATGTT
 AAAGTCTCAATTTCGGTTGAATGGCGCACATACTATCGAAGTCCGACTGGAGGCTGGAGAATGCGTT
 AGTTGATGACAATATTCTGGATGAACTTTTCCGACTGCAGCAAATTCTGAACGTTGCGCCTATTATT
 TATTGACGCATCTCATCAGGAGAAGGTTCAACGTTACTTATTGCAATTGTCAGCGACAGCTATGAA
 CCCATTGGCATGAGTACCGCATCAGTGGAAATGTTGTCGTATGCAAGGGCAATTAGAAGTTGTT
 GTATGAGCAAATGGTGAAGTCCAAAACAGTTGTTGGAGACGGTACGGAATAAGCGTCGTGGAAT
 TTTCCCAGGAGATACATAGTGTAAATGCTGTCAACAAAAGCCCTATGTTGGAGATAAAGGAGGGG
 CCATTGACCCACTCAAAGCTAAGGTTTTCTAAGTGGTATAGGGCATAACACCACGTTATTCTCT
 ATCTTATTCTATACATGCTGGTTACCATCTAGCTTCTCAAGCCGCAACCCCGCGGTGACCACCCCT
 GACAGGAGTAGCTAGCATTGACCACCCCTGACAGGATTAGCTAGCATATGAGCTGAGGATATCTACTGT
 GGGTACCCGGGATCCGTGTAGGCTGGAGCTGCTCGAAGTTCTATACTTCTAGAGAATAGGAACCTCGG
 AATAGGAACTAAGGAGGATATTCAAT

SEQ ID NO: 10 (example signal sequence for EPA carrier protein)

MKKIWIWALAG LVIASFASA

SEQ ID NO: 11 (example O1A *rfb* locus nucleotide sequence – O1A-EPA production strain
stGVXN4411 and stLMTB10217)

ATGACGAATTAAAAGCAGTTATTCTGTAGCGGGCTCGGGATGCATATGTTGCGCCACTAAGGCGAT
 ACCCAAAGAGATGCTACCAATCGTCGACAAGCCAATGATTGACTGACATTGTTGACGAGATTGTGGCTGCAG
 GGATCAAAGAAATCCTCTGGTAACTCACCGTCCAAGAACCGCGTCGAAAACCACTTCGACACCTCTTAT
 GAGTTAGAATCACTCCTGAGCAGCGCGTGAAGCGTCAACTGCTGGCGGAAGTACAGTCCATCTGTCGCC
 GGGCGTGAACATTATGAACGTCGCTCAGGGCGAACCTTAAAGGTTAGGCCACTCCATTGTCGCGCAG
 CTGCCATTGGTACAACCCATTGTCGTTACTGCCAGACGTTGTGATCGACGATGCCAGCGCCGACCCG
 CTACGTTACAACCTTGTGCCATGATTGACGTTCAACGAAACGGCCGAGCCAGGTGCTGGCAAAACG
 TATGCCGGGTGACCTCTGAAATACTCCGTATCCAGACTAAAGAGCCGCTGGACCGTGAGGGTAAAGTC
 GCCGCATTGTTGAATTATGAAACCGGATCAGCCGAGACGCTGGACTCAGACATCATGGCGTAGGT
 CGCTATGTGTTCTGCCGATATTGGCGGAACGTTGAACTGAGCTACTCAGCCTGGTGCATGGGACGTATTCA
 GCTGACTGATGCTATTGCGAGCTGGCAAAAACAATCCGTTGATGCAATGCTGATGACCGCGACAGTT
 ACGACTGCGGCAAAACGTTGAGCTATGCAAGGCTGTTGAGTATGGCTACGCAACCTGAAAGAAGGG
 GCGAAGTTCCGTAAAGGTATTGAGAAGCTGTTAAGCGAATAATGAAACCTGACCGGATGTAACGGTTGAT
 AAGAAAATTATAACGGCAGTGAACATTGCGAGCAAAAGTAATTGTCGAATCTTCTGCCGTTGTTTA
 TATAAACCATCAGAATAACACGAGTTAGCAGTAGGGTTTATTCAAGTTCCAGGATTTCTGTT
 CCAGAGCGGATTGTAAGACAATTAGCGTTGAATTTCGGGTTAGCGCAGTGGGTAACGCTCGTCAC
 ATCATAGGCATGCAGTGCCTGGTAGCTGTAAGCCAGGGCGGTAGCGTGCATTAAACCTCTATT
 AATCAAACGAGAGCCGCTTATTTCACAGCATGCTGAAGTAATGGAATAAAATAAGCTAGCGTGAAG
 ATACTGTTACTAGGGCGCAGGATTATTGTTCTGCTGAGTTGTCACATTATAATAACGCAAGGA
 TAGTGTGTTAATGTCGATAAAATTACGTAACGCCGAAACCTGGAATCACTGCTGATGTTCTGACTCTG
 AACGCTATGTTTGAACATGCGGATATTGCGATGCTGCAATGGCGGGATTGCTCAGCATCAG

CCGGATGCAGTGATGCACCTGGCTGCTGAAAGCCATGTGGATCGTCAATTACAGGCCCTGCGCATTAT
TGAAACCAATATTGGTGGTACTTATGCTTTGGAAGCGGCTCGCAATTACTGGTCTGCTTGTGCG
ACAAGAAAAATAGCTCCGTTTCATCATATTCTACTGACGAAGTCTATGGTGTGATTGCGCTACCTGAC
GAAGTAAATAAAAGAACAAATTACCCCTCTTACTGAGACGACAGCTACGCCCTAGTAGTCCTTATT
CGCATCAAAAGCATCCAGCGATCATTTAGTCCGCGCTGGAAACGTACCTATGGTTACCGACTATTGTGA
CTAAGTGTTCGAATAACTACGGTCCATTACCTTACCTTCCGGAAAATTGATTCCACTAGTAATTCTTAATGCT
CTGGAAGGTAAGGCATTACCTATTATGGCAAAGGGATCAAATTCTGTGACTGGCTGTATGTGAAGATCA
TGCCTGCGTTATACCGTAGTTACTGAAGGTCAAGCGGGTGAACACCTATAACATTGGCGACACAAG
AAAAGAAAAACATCGATGTTGTGCTGACTATTGTGATTGTTGGACGAGATAGTCCCAGAAAGAGAAATCT
TATCGTGAGCAAATTACTATGTTGCTGATGCCAGGGCATGATGCCGTATGCGATTGATGCTGAGAA
GATTGGTCGCGAATTGGGATGGAAACCACAGGAAACGTTGAGAGTGGGATTGCTAAAACGGTGAATGGT
ATTGGCTAATGCAAATGGGTGATAATGTGAAAAGTGGCTCATCAATCGGATTGACAGAACTAT
GAGGCCGCCAGTAATGAATATCCTCTTTGGCAAAACAGGGCAGGTAGGGAAACTACAGCGTGC
CTGGCACCTCTGGTAATTGATTGCTCTTGATGTTCACTGACTGATTACTGTGTTGATTTAGTAACCC
TGAAGGTGTGGCTGAAACAGTCAAAAGAATTGACCTGATGTTATTGTTAATGCTGCCCTCACACCGCAG
TAGATAAGGCTGAGTCAGAACCCGAAATTGACAATTACTCAATGCGACTAGCGTTGAATCAATTGCAAA
GCGCAAATGAAGTTGGGCTGGGTAATTCAACTCAACTGACTACGTATTCCGGAAATGGCGACAC
GCCATGGCTGGAGATGGATGCAACCGCACCGCTAAATGTTACGGTAAACCAAGTTAGCTGGAGAAAAG
CATTACAAGAGCATTGTGCGAAGCACCTAATTTCGTAACCGCTGGTCTATGCGATTGCTGGTTTGGAA
TTCGCCAAACGATGTTGCGTCTGGCAAAAGAGCGTGAAGAAACTAGCCGTTATTATGATCAGTTGGTGC
GCCAACAGGTGCTGAACTGCTGGTGTGATTGATGGCACATGCCATTGTCGCACTGAATAAACCGGATG
TCGCAGGCTTGTACCTGGTAGCCAGTGGTACCAACCTGGTACGATTATGCTGCCGTGGTTTGGAA
GAGGCCGCATGCGGATTCCCTTGCACTCAACAAGCTCAACGCGACTACCAACACTGCCCTAC
ACCAGCTCGTCCACATAACTCTCGCTTAATACAGAAAATTTCAGCAGAATTGCGCTGTATTGC
CTGACTGGCAGGGTGGTGTGAAACGATGCTCAACGAATTATTAGCACTACAGCAATTAAAGTGGGG
CATCTTGGTCGTGATGGTGGAGCAAGATGAATTAAAGGAATGATGAAATGAAAACGCTAAAGGTATT
TTTAGCGGGTGGTCTGGTACTCGTCTTATCTGTGACTATGGTGTCAAGTAAACAGCTATTACCTAT
ATGATAAAACCGATGATCTATTATCGCTTCTACACTGATGTTAGCGGGTATTGCGATATTCTGATT
AGTACGCCACAGGAACTCTCGTTCAACAACGCTGGTACGGTAGCCAGTGGGGCTGAATCTCA
GTACAAAGTGCAACCGAGTCCGGATGGTCTTGCAGGCAATTATCGGTGAAGAGTTATTGGTGGTG
ATGATTGTGCTTGGTACTTGGTGTGAAATATCTACGGTACGACCTGCCTAAGTTAATGGATGCCGCT
GTTAACAAAGAAAGTGGTCAACGGTATTGCTATCACGTTAATGATCCTGAACGCTATGGTGTGTTGA
GTTGATAAAACGGTACGGCGATCACGCTGGAAAGAAAACGCTACAACCAAAAGTAATTATGCCGTA
CCGGCTTATTTATGATAACGACGTTGCGAAATGGGAAAATCTAACGCCCTGCCCGCGTGAA
CTGGAAATTACCGATATTAACCGTATCTATATGGAACAAGGGCGTTATCTGTGCCATGATGGGGCGTG
TTATGCGTGGTAGACACGGGACACATCAGAGCCTGATTGAGGCAAGCAACTTATTGCAACAATTGAAG
AGCGTCAGGGCTGAAAGTTCTGCCGGAAGAAAATTGCTTACCGTAAAGGGTTGATGCTGAGCAG
GTGAAAGTATTAGCTGAAACCTCTGAAAAAAATGCTTATGGTCACTGCTGAAATGATTAAAGGTTA
TTAATAAAATGAACGTAATTAAACAGAAATTCTGATGTTACTGATTTGAAACGAAAGTTTGGTGT
GAGCGTGGTTCTTTGAGAGCTTAACCGAAGGTTTGGAGGAGCTGTAGGCCGAAAGTTGAATT
TGTCAGGATAACCATTGAAAGTCTAGTAAAGGTGTTACGCCGCTGCATTATCAGTTGAAACCTTATG
CACAGGAAAATTGGTGCCTGCGTGTGGTGAAGTTTGACGTAGCTGTTGATATTGCTAAATCGTCA
TCGACTTGGCAAATGGGTGGGGTGAATTATCTGCTGAGAATAAGCGGCAATTGTTGAGGG
ATTGACATGGTTTGTGAGTGAGACGGCGAGTTTGATAAGACGACAAATTATCATC
CTCAGAGTGTAGAGGAATAAAATGGGATGATCCAAGCATCAATATTGATGCCAGTCGATTCAAGTG
CTGCTATCAGCTAAAGATAATAAGCATCCCATTAACAAAGATTGAAATGTATAGTTAAGATCAGATAA
ATCTGGAAGGGTTGCAAAATTGAATAAAATAGTGAGCAAAGTGAATAAGGAACGTAATCCACAAATGCT
GGCTATGATGATTACTCAGATAGCTTATGTTGACCAATTACTGAGTTATCTGTTAAAAC
ACTGGGGTGGCACAGTTGGTAATTGCTTAATACTATCAATCGTGCATATTACAGATTATAACGG
ATTATGGTTTCTTGTGCAAGCTGAATTAAACAGGAATTATGATGTTGCTTGTAAATAGGAATA
AAATCTTGCTGTGCAAGCTGAATTAAACAGGAATTATGATGTTGCTTGTAAATAGGAATA
CTTCCAAACCAATGGTTTCCAAGGTATGAAAGAAAATTAAACATAGCCCTTCTAATGTTAT
AGATGCGCCGCGTGTACTGTTACTGTTATCTATGAGGAATAGCGAGGATTACAAAAGCACTTTAGT
ACAGTCACCTCATTAGTAATTCTGCGATTGGATTAAATATTTATATTGAAATATCAATATTATT

TTCCGGAAAAAAATTATTAAGGTAATTAAAAGAAGGTAGGATTTTCTGCATCACTTTATTCT
GTTATTCTCAATAATAGTGGCATTTCATTAGGGATTACTAATCCTGTATTGTTGGTGTATATGC
CGCCGCTGAAAAGATAGTCAAGGCCGTATTGTCGCTATTACACCCTGACGCAAGCTATATATCCTATA
ATTGTCGTAAGTTTCACTATCCGTATTGACGGCATTGAGGCAGCAAAAAAACTGGTATACCAATTATA
ATTTAGCATTATAGCTGTTACGTTGCAATTACCTTACCTGTTGCAATCGACTATCTTAATTTC
AAAAGAAACAATTTTAGGTCAAATATTAAAGTGCATGGATCTTTGGTGTCTTAATAATGTATTG
GCATTCAAGATATTGAGTGCATCAGGAAGAAGTAAAATATAGTAGGATGGTATTGCAATCAGCGCTTATA
ACATTACTTTGATTACTTATTATGCAAGTGTGTTGCAACGCCACTGGAGTGGCATGTGCAATATTGGG
TGAAATGTTCTTATCAATATTGTTACTTAAGCGATAAAAAAAATAATTAAAGGAATAGTTATGAAGAAGT
TATTATTAGTGGTCTGGTACTAGGCCTGAAGCAATAAGATGGCCTCATCATTGAATTATTAAAAAGAT
TGTAGATTGCAATATAAAATATGTGTGACAGGCCAACATAAGAGATGCTTGATCAAGTTATGCAAGTATT
TGATGTTAACACTGATTATAATTACGGATTATGCAAGCCTGGCAAACATTAGTATCTATAGCAACAAATA
TACTCTCACGGTTAAGTGAAGTTAATTATAGAAAAGCCAGATATTACTTGTGATGGGATACAACG
ACTACCCTGCTGCTACTTAGCTGGTATTACCAACAAATAAAAGTTGTCATGTGGAAGCAGGATTAAG
AACAGGGGATATTACTCTCCTGGCCTGAAGAGGGCAATCGTAAAGTTACAGGGCATTAGCATGTATT
ATTCGCCCAACAGAGAGTCAAAGATAACTCCTGAGGGAGGGGTCAAAGTAAATAATATTGTA
ACGGTAATACCGTCATCGACTCTTATTGCAAAAGATACTAGATAATGACCTAATATAAAGAA
CGCTTACATAATAAAATTAAATTCTGATAAAAGCCGACGAGTAGTACTTATAACAGGTATCGAAGAG
AAAATTCTGGGAAAGGTTGAAGATATATGCTTGCAATAAGGAATTAGCTTCATTATCCTAAATGTA
GATTATTATTCAGGTCATCTTAATCCAATGTAATGAAACCAGTACATGTATATTAGATAATATG
TAATATTACCTTATTGAGCCCTTGATTATTGCTTTGTTATTAAATGAATGAGTCATATTAAATAT
TGACTGATTCAAGGGGGATAACAAGAAGAAGCGCTTCGTAGGTAACCGGTTTGGTTATGCGTGTACT
ACTGAACGCCCTGAGGCGGTGAGGCTGGTACTGTTGTATTAGTGGGACTTCTAAGATAAAAATAGTAA
TAAAGTAACGGAGCTATTAAACAATGCTGATATCTACAATGCTATGCTCTGTTACATAATCCATATGGCG
ATGGAACAGCTGCTCAAAAATTCTTAATGTGCTGCCAACAGAGCTAATTAAATTAAAGCTAAAATATGT
TATTAATTATTGCTGATTATCCAAACGAAATGAATATGCGCGAGGGAGCTATGCAACGAATAGATGCGATA
GAECTCTCATCGAGATCGCAAGCGAGTGTATTGAATATTCAAAAGCATCTAGTCGCTCAAA
TAGTCCTTAATAATGTATAGTGAAGGATCTAAATGCAATTATTACAGAAACATCATAAAACAGTACA
TGCAAAATCAACAATATATGTCATTCTGTTATAATTAAAGGTATAACGCTCATTGATCT
AAAAAACAAATCTTGTATACATGGTGTACCGGAAGAACCTTGGCAGATAATAAAATTACTTAG
TAAAGTATATAACATGGTGGAAAAAAAGGTGCTCTGGATGCAAAATAACACGTCACTACAGAAA
TGCAAAACACTATGAAGCAAAATGGAGTAAACTGGCTGAAAGTCATAAGTGCTCCGATTTGAA
TATAAAATATAACCCAATCGAAACAAATGGACAGAAAATAACGAAGTATCTATCTGGAGGATT
ACAAACATGGCAAAATATGATAAAATGATTCAAGTTGATGACACAGTGTAAACAAATGAAGCAGGTA
AGTATGAATTCAACCTTTCATCCCACAGAGTAACTTGGAGGGTTATAGATAATATTGTTAAATT
CATATAATCAATGCTAATGCATCTACGCTATCACGTGATGAAGTAATTCCCTTCTAAAGAATGT
TGGTTTGATTGCGCGATGATATAATAGTAAACAGAGTGTGCGTGCCTACAAAATTGGTGAATATTAG
AGTGTGGTGTGTTCCAGTTGCTCTCCCCACTTATAGGTGATTGATTCGATGGGATATCAATACATT
ACTACAGAGGAAATGGCTAACAGAAGTATAAGTTGTTGATCTGAAAAAATGGCTGCACATAATTACA
AATTGACTTCTTATCAGAAGAGAACCTACAAGGCACAGAAAAGAACCTATTGCTCAACTGTGCTGAATT
TTTACATATATAAAATTATGTAAGCATACTGCGGGTCAGGTAAATTGATGCGTATCAAATATAAGATAAC
GGTTATATATTGTTCTATTGTTGAGCTACTTAGTTACTCAAATCTGACTACTTCC
TGCTGATTTCTGCCATATACAGAAATACGATGGGACATACGGAGAAATCAATAATTGAGCCTGCCT
TTTATATTAACACGGTGTTCATTATTAATTCCCTATATATTGCAATGTAGTTG
TTATGTTAAGTGGAAAATAAAATATGCAAGAAAATAATTAAAGATAGTTATATATTGTTCTGTA
TGTATATGTATCTTATGTTGTTGATGAAATGACTCAATTGCGCATAGCAATTGCAAGTC
GCTATGTCGTTTATTATTACTTTATAAAATTGTTAAACATGCACTGCCATGGATGGTGTGGCT
ATTGTTTCATTACAGGCCCTGCTTTATTGTCATTATTATACAGTTAGGAGGTTATTAAAT
AGTAATTATAGGGTTGTAATATGTATGAGCTTTAAACGTGTATGCAAGTACATTC
CAAATGAAAAAATAGTAAATTATTATAGTATTTCATCATCATTAGACAATAGAAATGATTGGCAATA
TTCAACCTGAATAATATAATTTCATTCAATATTATTGATCTTTATCTAGCGATATATAAAATT
AAATGATAATGAGGCGAAGTTATTAAAGTATGTCATGCAATTGTCAGGAATTAGCCTTGTATT
TGGCTAGTGGAGTCCCGGTATTGCTTATCGAACTGCAAGAGTTGCTGCAATT
GTATTAACTCTTCGCATATAAAATAATGCGTATTGCAAGTCATTAGTTATC
AGGCTTAATGTTGTTATAACACTAAGGGCTGATCAATAGTTGGTCAAGGATTATAAAATGAATGTTGCT

ATTTGTTGTCTACGTATAATGGCGAAAATATTAGAGGAACAATGGATTCTTGCTGCTCAAAGTTA
TCAGGATTTGTAGTGTATATCCGTGATGACGGATCATCTGATAGAACTGTAATATAATAAACCAATACG
TAATGAAAGATAACAGATTATTAAACGTGGTAATTCAAGAAAATCTGGTTGCTGCTTCGTTATTAAAT
TTATTAAAGAAATGCTTCAGCCGATATTATGTTGACCAAGATGATTATTGGCTTCCGAATAAATT
ACAGCGTGCTGGATTATTTCGGCTATTGATCCTTACAACCTACCTGTATCATTGCGATCTAAGCG
TTGTTGATGAAAACCTTAATATTACAAAATTCACTTTGCAGCATCAGAAAATGTCAGCGTATGATTCA
ATGAGAAAAATAATCTTCATACAAAATTGTTGTTGGTTGTTCATGTGCTGTTAATGCTTCACTTGC
GGAATTGTTCTTCGCGAATTGGAGAGCAGCATGTAACGTTGACTGACTGGTGGTTAGCCG
TGACTGCAAAACTTTGGTCAACCATTGATAATACTCAAACGATTCTTATCGACAACATCAGGGC
AATGTTAGGTGCAAAATCATCAGGTATGATGCGTTATTGATTAGGATAATGGGCAAGGGATTTC
GCGAGTAGTATTTAGAAAAAAAGTTGCGCAAATAAGCTCTTTAGATGTCTATGATAAAGATT
TAAATCTTGAGCAAAAAAAATCTATCAGGCTTGTATTGAGGGCCTAAAGAGAACTCTCAATTGCTGAC
CTTTAAAATGTTCTATCATGGTAGCTATGCAAGGTTAAACGTAATCTGCCTTAATATATTCACT
TCTTACACAAAAAGAAGATAGTGTATCCTTATGAAAAAAATTGCTATTATCGGTACTGTTGGCATA
CAGCATCATATGGCGGATTGAAACATTAGTTGAAAATTAAACAGATAACAATTCTCGGGAGTTGAATAT
AATGTTTTGTTCATGTTCACTACAAATCCCACCAAAAAAACATAATGGGGCCGTTAATTATAT
TCCGCTTAAAGCCAATGGATGGCAGAGCATTGCGTATGACATAATTGTTAGCATATTCTATTG
AGCCTGATGTGATTCTGATTAGGGTTCTGGTTGTCATTGCTTCTCAAACTCTAACACCGC
GCTAAGTTTACTAATATTGATGGCCTGGAATGGCGAAGAGATAATGGAATTCAAAGTGAACCGTT
CTTAAATTTCAAGAAAAATCGCAGTTCAATATTGCGATGTCGTTATTACGGATAATGAGGCAATTCTG
AGTACGTTTAAACGAGTATAAAAGATAGCCGAGTTATTGCTATGGAGGGATCATGCATGGTAAAT
ACTGAGGATGATTACAACAAGAAATTATAAAAGCATTACTACCTTCTGTATGTCGATCGAACCGA
AAACAATGTAGAATTAATTAAACATTTCAGCTAAAGCTAAATATAAAATAAAATTGAAATTGATG
ATGGCAGCGAGTTGAAAGAAACTAGGCTGCATTACTAACTATCAAATATTGAAATGATTGATCCG
ATTATGATCTCAACAATTATTCACTACGAAATAATTGATAGGATATACATGGTCATTGGCTGG
AGGAACAAACCTCTTAGTCGAGGCAATGCAATTAGTAAACCTATATTGATATGATTGTAAGTTA
ATAGGTACACTACTGAAAATGAAGCATGTTATTCTAATGAATCTGACCTCGCAGAGAAAATCATAATG
CATTGAGCTATCATTAGGTGTCCTGGCACGAAATGAAAGAAATTGCTAACAGAAATACACTGGAG
ACGAATAGCAGAAATGTAGGATGCTATTAACTCTGTTAAACTCAAATCTTACATAATATGGCAT
GACTATAAGCGCATTAAATTGTTCAAGCCGCTCGCGGTGACCACCCCTGACAGGGGATCCGTG
GCTGGAGCTGCTTCGAAGTTCCTATACTTTCTAGAGAATAGGAACCTCGGAATAGGAACTAAGGAG
TCATATGGATAAAGCCGTAAGCATATAAGCATGGATAAGCTATTATACTTAAAGTACTTGTATA
TATTGCGAACATTCCAGGCCGAGCATTAGCGCGGTGATCACACCTGACAGGAGTATGTAATG
GCAACAGATCGCGTAGTCGGTATGGCAGTGATGGGACGCAACCTTGCCTAACATCGAAAGCG
ATACCGTCTCTATTTCACCGTCCCGTGAGAAGACGGAAGAAGTGTGATTGCCGAAATCCAG
CTGGTCTCTACTATACGGTGAAGAGTTGCGAATCTGAAACGCCTCGCATCTGTTAATGGT
GAAAGCAGGTGCAAGGCACGGATGCTGATTGATTCCCTCAAACCATATCTGATAAAGGAGAC
ATCATCA TTGATGGTGGTAACACCTCTTCCAGGACACTATTGCTGTAATGAGCTTCAGCAG
GAAAGAAGCCTATGAATTGGTAGCAGCCTGACCAAATGCCCGTAGCTGAAGACGGTGAACCAT
GCGTTACCTATATTGGTGGCATGGCGCAGGTCACTATGTAAGATGGTCAACACGGTATTGA
GATATGCAGCTGATTGCTGAAGCCTATTCTCTGTTAAAGGTGGCCTGAAACCTCACCA
GCAGACCTTACCGAGTGAATAACGGTGAACCTGAGCAGTTACCTGATCGACATCACCA
CCAAAAAAGATGAAGACGTAACCTGGTGTGATCTGGATGAAAGCGCTAACAAAGGTACCG
AAATGGACCAGCCAGAGCGCGCTGGATCTGGCGAACCGCTGTCGCTGATTACCGAGTCTG
TTATATCTCTCTGAAAGATCAGCGTGTGCGCATCTAAAGTTCTCTGGTCCGCAAGCACAG
CAGCGACAAGGCTGAGTCATCGAAAAGTCTCGTGTGCGCTGTCTGGCAAATCGTT
CAGGGCTTCTCTCAGCTCGTGTGCGTGAAGAGTACAACACTGGGATCTGAACTACGG
GATTTCGCGTGTGGCTGCATCATCCGTGCGCAGTCTGCTGAGAAAATCACC
CACAGATCGCTAACCTGTTGCTGGCTCCGTACTTCAAGCAAATTGCCGATGACT
GATGTCGTTGCTTATGAGTACAGAACGGTATTCCGGTCCGACCTCTCCG
CAGCTACCGTGTGCTGTCTGCCTCGAACCTGATCCAGGCACAGCGT
GACTATTGGTGCCTACTTGGCTGGATTAA
ATAAGCGTATCGATAAAGAAGGTGTGTTCCATACCGAATGGCTGGATTAA

SEQ ID NO: 12 (example O2 *rbf* locus nucleotide sequence – O2-EPA production strain stGVXN4906)

ATGACGAATTAAAAGCAGTTATTCTGTAGCGGGCTCGGGATGCATATGTTGCCACTAAGGCGAT
ACCCAAAGAGATGCTACCAATCGTGACAAGCCAATGATTCACTGACGAGATGTTGCTGAG
GGATCAAAGAAATCCTCTGGTAACTCACCGTCCAAGAACGCGTCGAAAACCACCTCGACACCTTTAT
GAGTTAGAATCACTCCTTGAGCAGCGCTGAAGCGTCAACTGCTGGCGAAGTACAGTCCATCTGTCCGCC
GGCGTGAACATTATGAACGTGCGTCAGGGCAACCTTAGGTTAGGCCACTCCATTGTCGCGAC
CTGCCATTGGTGACAACCCATTGTCGTTACTGCCAGACGTTGATCGACGATGCCAGCGCCGACCCG
CTACGTTACAACCTTGCTGCCATGATTGACGTTCAACGAAACGGGCCGAGCCAGGTGCTGGCAAAACG
TATGCCGGGTGACCTCTGTAAATACTCCGTATCCAGACTAAAGAGCCGCTGGACCGTGAGGGTAAAGTCA
GCCGATTGTAATTATCGAAAACCGGATCAGCGCAGACGCTGGACTCAGACATCATGGCGTAGGT
CGCTATGTGTTCTGCCGATATTGGCCGAACTGGAACGTACTCAGCCTGGTGCATGGGACGTATTCA
GCTGACTGATGCTATTGCGAGCTGGCAAAAAAACATCCGTTGATGCAATGCTGATGACCGGCACAGTT
ACGACTGCGGCAAAAAATGGCTATATGAGCGTTGAGCTGGCAAGTAAATGAAACCTGACCGGATGTAACGGTTGAT
GCGAAGTCCGTAAAGGTATTGAGAAGCTGTTAGCGAATAATGAAACCTGACCGGATGTAACGGTTGAT
AAGAAAATTATAACGGCAGTGAAAATTGCGAGCAAAAGTAATTGTCGAAATCTTCTGCCGTTTTA
TATAAACCATCAGAATAACACGAGTTAGCAGTAGGGTTTATTCAAAGTTTCCAGGATTTCTGT
CCAGAGCGGATTGTAAGACAATTAGCGTTGAATTTCGGGTTAGCGCAGTGGGTAACGCTCGTCAC
ATCATAGGCATGCAGTGCCTGGTAGCTGTAAGCCAGGGCGGTAGCGTCATTAATACCTCTATT
AATCAAACGTGAGAGCCGCTTATTTCACAGCATGCTCTGAAGTAATATGGAATAAAATTAGTAAAATACCT
GTTACTGGTGGCGCAGGATTATTGGTTAGCTGTCACATTATAAAATACGCAGGATAGTGT
TGTAAATGTCGATAAAATTACGTACGCCGAAACCGGGAAACTCTGCTGATGTTCTGATTCTGAAACGCT
ATGTTTTGAAACATGCGGATATTGCGATGCACCTGCAATTGGCACGGATTGCTCAGCATCAGCCGGAT
GCAGTGATGCACCTGGCTGTAAGCCATGTTGACCGTCAATTACAGGCCCTGCGGCATTATTGAAAC
CAATATTGTTGTTACTTATGCTTTGGAAAGCCGCTCGAATTACTGGTCTGCTCTGATAGCGACAAGA
AAAATAGCTCCGTTTCATCATATTCTACTGACGAAGTCTATGGTATTGCTCATCCAGATGAAGTA
AATAATACAGAAGAATTACCCATTACTGAGACGACAGCTACCGCACAAGCAGCCATTACCGC
CAAAGCATCCAGCGATATTAGTCGCGCATGGAAACGTACGTATGGTTACCGACCATTGACTAATT
GCTCGAACAACTATGGTCGTACTTCCCAGGGAAAGCTTATTCCATTGGTATTCTAATGCACTGGAA
GGTAAGGCATTACCTATTATGGCAAGGGGATCAAATTGCGACTGGTTGATGTAGAGGATCATGCTCG
TGCCTTATATACCGCTGTAACCGAAGGTAAAGCGGGTAAACCTTATAACATTGGCGGACACAACGAAAAGA
AAAACATCGATGTTGCTGACTATTGTTGATGGGATGAGATTGACCGAAAGAGAAAATCTATCGT
GAGCAAATTACTTATGTTGCTGATGCCAGGGCATGATGCCGTTATGCAATTGATGCCATAAAATTAG
CCGCGAATTGGGCTGAAACACAGGAAACGTTGAGAGCGGGATTGCAAAACGGTGGATGGTATCTGG
CTAATACAATTGGTTGAGAATGTAAGGAAAGCGGTGCTTATCAGTCATGGATCGAACAAACTATGAGGC
CGTCAATGAATATCTGCTTTGGCAAACAGGGCAGGTGGGTTGGGAACTGCAAGCGTCTGGCG
CCGCTGGTAATCTGATGCTCTGATGTTCACTCCACTAATTATTGAGGATTTGCAACCCCCGAAGG
TGTGGCAGAAACCGTCAAAATTGCTGACGTTATGTTAATGCTGCTGCTCACACTGCACTAGATA
AAGCAGAATCAGAACCGGATTTCGACAATTACTTAACCGACAAGCGTCAAGCGATTGCAAAAGCTGCT
AATGAAGTCGGGCCTGGTTATACACTACTCTACTGATTGTTCCACGGCAGTGGTGACGCGCCATG
GCTGGAAACGGATGCAACAGCACCGCTAAATGTTACGGTAAACAAAATTAGCTGGGAAAAGGCATTAC
AAGAACATTGCGCAAAGCATCTTATTCGTAACAGCTGGTATACTGCTGGTAAAGGAAATAACTTGCT
AAAACGATGTTGCGTTGGCAAAGAACGCGAAGAATGGCTGTGATAAACGATCAGTTGGCGACCAAC
AGGTGCTGAATTGCTGGCTGATTGACCGCTCATGCCATTGCGTGGCATTAAAAACAGAACGCTG
GCTTGTACCATCTGGTAGCAAGTGGCACAACAACCTGGCACGATTATGCTGCGTGGTTTGAAGAGGCG
CGCAAAGCAGGGATTAATCTGCACTTAACAAACTTAACGCCGTGCAACAAACGGCTATCCCACACCAGC
CCGTCGACCCATAACTCTGCCCTCAATACAGAAAAGTTCACTGAGAACACTTGGCCTGCTTGCCTGACT
GGCAGGTGGCGTGAACGTATGCTCAACGAATTATTACGACTACGGCAATTAAACAAATTGTCATCT
CGCTCATGATGCCAGAGCGGGATGAATTAAAGGAATGGTAAACACGCGTAAAGGTATTATCTGG
CTGGTGGTCCGGCACTCGTCTTATCCTGTGACGATGGCAGTGGTAAACAAATTGCTGCCATTATGAT
AAGCCGATGATTATTATCCGCTTCAACGTTATGTTAGCGGGTATTGCGATAATTCTTATTAGTAC
GCCACAGGATACACCGCTTCCAACAATTATGGGGGACCGGAGCCAGTGGGGTCTTAATCTACAGTATA
AAAGTACAACCGAGTCCGGATGGCCTGGCGAAGCGTTATTATGGCGAAGACTTATTGGTGGTGTGATGAT
TGTGCACTCGTACTTGGCGATAATCTTATGGACAGCACTGCGAAATTGATGGAAGCTGCTGTTAA

CAAAGAAAGCGGTGCAACGGTATTCGTTATCACGTTAATGATCCTGAACGCTATGGTGTGGAGTTG
ATAATAACGGTACGGCAATTAGCCTGGAGAAAAACCGCTGGAGCCAAAAGCAACTATGCGGTTACTGGG
CTTATTCTATGACAATGACGTTGGAAATGGCTAAAACCTAACGCTTCTGCCGTGGCAACTGG
AATTACCGATATTAACCGTATTATGGAACAGGACGTTGTAGCCATGATGGGCGTGGCTATG
CATGGTGGATACAGGGACGCATCAAAGCCTTATTGAAGCAAGTAACCTCATGCAACAATTGAAGAGCGT
CAGGGATTAAAGGTATCTGCCGGAAAGAGATGCTTACCGTAAGGGTTATTGATGCCAGCAGGTGAA
AGTATTAGCCGAACCGCTATCAAGAATCAATATGGTCAATATTGCTGAAATGATCAGCGAATAGTATA
TGGGAACTCAATGATGGATATTAAATTATCTTGCRAAAACATGGGATGAGCGCGGTGATTAATTG
CTCTGAAGAGCAACGAAATATACCTTCAAGTCAAAGAATATATTACATACTTGAGACTCTTAATGGA
GTAAGACGCCATTATCGCACAAGGTTACTCGTAGTTGCTATTGAGTCAGGGAGCTTGTAAATT
TCATCTGGATAATGGTAAAGAAACAAAGCAGGTGGAACCTAATGATCCAACAATTGCGTGTGATAGAAC
CCTATATATGGCATGAAATGTATGATTTAGTGTGATTGTGCTGTTGTAATTGCGGATGATTCTAT
AAAGAGTCTGATTATATCCGCAATTATGATGATTTATTAGAAGAGTAAATTCAATTGAGAATTCAAGC
TAAGTGACGTCAGACAACTCAATTGGTGTGAGAACACTATCTGGCAGTTGTGATGAACTAAAGGT
GCTGTAATTGGTAAATAATTGCAACATCTGTGCAAATACCTTAATTGAAAATAACGTTGTAATTGGTAAACA
TGTACAGTCAAAAGCGGTGTATATTGGGATGGCTTAAAGTCTATCTGATGAAATTGCAAAACAATA
CGCAAAGGAGCATCAATTGGTGTGAGAACACTATCTGGCAGGAAATTGAAATTGGTGAAGAACATCGT
TGGTGCAGGGAGTGTGTAACCAAAATGTACCGCCATGCGCAATAGTAGTAGGTAATCCAGCTCGATT
TTAAATGGTAGAGGATAATGAATAAAATTGATTTTAGATCTTGTGCAATTACCAGCGACAGCACAA
AGAATTAGTCTCGGTTAGTAGGGTGTAGATTCTGGTGTATATCATGGCGAAGAACTTGAGCAGT
TCGAGAAAGAGTCGAGAAACTGTGGAGTTAAGTATTGCAATTGGTGTAGCAAATGCCCTGATGCGTT
ATACTAGTATTGAGGGCATGGAAAGAACCTGGCTATCTGAAAGACGGTGTGAGGTTAGTACCGGCAAA
TACATATATTGCTCTATTCTGCTATAACAGAGAACAAACTGTGCTCTGGTGTGAAACCAGATATAG
AAACTTATAATATTAAATTCTGCTTAATTGAAAATTACATTACGGAAAAACTAAAGCAATTACCGGTT
CACTTATATTGCTATTGTCAATATGCCAGAAAATTAGTCAATGCCAGAAAATAATCTGTTGATTCT
TGAAGATTGTGACAAGCACATGGTCAATACGTGATGGTCGAAAGCTGGAGCTGGGGATGCTGAG
GATTAGTTTATCCAGGAAAAACCTGGAGCTTGGGGATGGGGAGCTGTTACTACAAATAATGCA
GAATTATCCTCAACTATAAAAGCTTGCAGAAATTGGTCACATAAGAAAATGAAAATTATCAGGG
ATTGAATAGTGCATTGGATGAACTGCAAGCAGCCTTATTGCGTGTAAAATCCATACATTACCGGAAAGATA
CTGCGATTGGCAAAGGATTGCTGAAAATATTCGTGAAATAAAAACCTGCGATTACGTTACAGTG
TACGAAGGCCAAGGTGCGCATGTTGCATTATTGAGTAAGAATCGTAATCGTGAAGAACCTCAGTC
ATACTTATTAGAGAAGGGTATCAAACCTTAATTCACTATCCATTACCAACCCATAAGCAGCAAGCATATC
AAAATATGCTAGCCTAGCCTCCAATTACTGAGCAAATTGATGAAAGTCATTCTTACCTATAAGT
CCGGTAATGAGTGAAGATGATGTCAATTGTAATCAAATGGTCAATGATTACAAGTAATGAAAAATT
CTTCAGGTAACTATATTACCGCTATCTACATTAAATGATTGCGGGTTTATCATCGGTAAAGGT
AGTAGCAATTATACAGGCCATAGGGTAGCAATGCTGGCAAGTGCAGGTTAATCACAATAGTTG
CAGGTACTACCTCTGCACCTGTAAGCACAGGCCATTGCGTGTGAGTACTGCGGAAAATTGGCAAGAAGGACAA
GAAGCATGCGGCCATGGTGGCGCATGCTTAAGGGTACTCTGTTTATTGCTTATTACCGGTT
TGTTATTATATTGCAAAATATTAGTAGTTGAGTTACTTTAGCGATGGACAATACACATGGTTAATCATT
TCGCATGTTGATATTGCAATTCTCATTATAAAATACATTGATCGCTTCAGTTAAATGGTCAACAATT
TATAAGCAATATATTGGTGGGATGTTCTGTATTCTACTATGTTATGATTGTTGATTG
AGCTTATAATCTTAAAGGTGCATTGATTGCCACAGCTATAAAATAGTGTATTGCTGGTCTGTATTGGTT
TATTGCTCAATAATCTGGTTAGATTAAATATTGGGGTAAAACGGATAAAGACAAAATTATA
AAAATTATTCAATTACTCTGATGGCTGGTTCTGTTATCTCCATGCTACAGCATTGATGTTGATTAG
AAAATATTGATTGCTAAAAGTGGTGGGAGGATGCAAGGGCAATGGCAGGCCGTATGAAAGATATGAGG
TTTATCTGGTGTGACAATTGCTTGTCAACATATTCTACCAAGATTGACAATTATAAAAACAAGT
TTCCTTATAAAAAGAAGTAAATAGTACTATATTACATAATATCTATTACTCATGGCGTTGAG
TATCTATTATCCCGATTGGTAAATAACAGTTTATTACTGAAACAGTTGCTCAGCTCGTGAATTAT
TTTATTACAACTTATAGGGGATGTAATAAAATTGCTGGTTCTTATGCAATTACCTCTCAAAGTCAG
GGGCAACTAAACTATTACATCAGTCAGAAGTGTATTCTCATGCTTATTACCAACCTATA
TGTTGAAATTATGGAGTACATGGTCTAACATAAGTTATGTCATTACATATAGTTTATATTGTTG
CATTGTTACTAATTATTAAATGTTAGAAGAAATAATTAAAACAGAGGTTGAATTGAAAATAAT
TATACTGCTTAGGATTGGCAGGGCTGGTGGTAAAGAGTTCTTCAAGCTGGCAACTGAATTGATGA
ATTATGGACATGATGTAAGTTGTTGTCAGATAATAGAAACTAATCCATATTGCTACCAACAGCAAA

ATTGTCACTGAGTAAATCTAGTCAAAACCGTGTAAAAAATATTGAGAATCATTAAAATTACTATAATCTGT
GCGTAATGCATAGAGTTAACATCCTGATGCTGTTGCTAGTTCTATTGACTGCCTATCTGTGCGCAT
TATTACCAATCACCGTCGTAAGAAATATTATATTACAGGCGTATGAAGTTAATTGGTTGATAATATA
ATATGAAATTAAAGCCGGTTAACATATTATTACAGGCTTAAAGAAATACTAAATAGTCCTAATTGCT
TCCTCATAAACATGATGATTAGGAGTAGTCCTGCAAGGAGTAGATTAAACGTTCTATCCGAAAC
CATCAAATAGGTTAAATGGTCACACATCAATAGGGATTATTGGTAGAAAGAGCACAAGGAAC
AGCAGAAATTATTCAGTATTGTTCACTGGAAATAAGCTGGATTATAATCAATATTGCGATCTATCT
TGAAGAAGTTGATAAGCAGCTTAACTGCTGCCGGTTTCAGGTTAATTGGTTGCGATTACTCTGATT
TAGAATTGGCATCCTTATCGAAGCAATGACATCATGATTGCTGTTGGTTATTGAAGATGGCGTTTC
CATTATCCTGTCGAATCAATGGCTGTTGTTATTCAAAATTATGCCCACTTACTGAAAC
TAACAGTGTACTAAATTAGTCAAGTTGATGCTGCAAACCTGGTGAAGCAATTAAATCTTGCTCAATC
TTGACCTAGAAGAAAAAGCAAAGAAATCCAATCTAATATTCTGTTGAATAAAATGACTGGAAATT
GTTGGTAAACTTCAATAGTTATTGTTAGATGCAAATAAATAGTATACGTTGATGGGGAAATATGAAT
ATTGTTAAACTGATATTCCAGATCTGATGTTCTGAAACCAAAAGTGTGTTAGTGAACGCGGCTTTT
TATGGAGAGTTATAATCAGATTGAGAAGGCAATAGGAAGGCACGTAATTGGTCAAGGATAATC
ATTCAAAATCTAGAAAGGCGTACTACGGTTGCATTATCAATTAGCACCGTATGCACAGGCTAAATT
GTTGATGTTGAGGTATTGATGTTGCTGTTGATCTTAGAAAAAATTACCAACGTTCAAAA
ATGGTTGGAATAACCTTCCGCAGAAAATAACGACAATTATGGGATACCGAAGGATTGCTCATGGTT
TCTGGTGACCAGTGATGAAGCTGAGTCATTATAAGACAACTAACACTATGCTCTGGTCATCAGCAA
GCAATTATTACAATGATCCTATTAAACATCGATTGGCCTTCTGCACTAGTGTCTGTCATTATCACA
AAAAGATCAAGAACGAAAATTATTTCAAGGATTATGGACAGTGAACCTGTTCTAATAAAGTGTGCCACCT
ATCCGTCGAAGGATAGGTGGTTGCTTATATTGGTGAAGTATGTTGATAATGACAGAAATAGTCGGA
AATATAAACACGATAAAAGCTTAAAGTTTATCTACTTATTGGTATTTACACTTATTGAGGCTT
ATTATCGCAAATACCCAGTTTGGGGCGAAGTAGAGACTATGATAATTATACAGATTTCTGGTAA
AGAAGGGAGGGGGTCTGAAATTATTTATCGGGATTGATGTTATAACGACCACTGAAACTATCA
TTTTATAATTAAACATGTTCTTTTATAAAGGCAAGGTTCTCGCTAACTATTGCGTAATTGGTCA
GGCTTGACCTATTCTTATTATGCAAGCGTTGCACTTGGGTTTAGATTACTCAATTGAGAAA
TGGTCTATGATTCCATTAAATGTTCCGTTACTATTATTATAAAATAACGACTTATTGGT
TCTCGGTATTATGTCGAATTGCAACTCATGGTCTGTTGCCCTTGTCTTATATCCTTGTCTAT
TCAACAAAATAAGACGCCGGTTATTGTTGAGTATTCTGTTGATGTTGCTCATCAGGAGAAGG
AAAAGAGATCATATTGTTATAAGAAATTGGAGTGGGACAAAAAATAGGAAATGAAGCTGGTAAATT
TAATAAATTCAATTATCCCTACCGCTATTCTGGTTATTAGTTACATATCAAGCATTGGAAATGAA
AGGAGAAATTAAAGGCTTCTTGTATGGGTGATGCAATACGTGACTTTAGCCTTCTCTTAC
TGTATGGCTTCCGTTGGAAATGATTCTTCTGCTAACCATGGGTTGTTATAAGCAA
AAAAGAATTATTATTTATTGGCAAAAGTGTAAATTGTTATGTATCTAACACTATTATCATATGGTC
TTGGAGTGATTAATGTGTAAGGCTAAGGTGTTGGCTATAATTGTTACTACACCCGAAATTTCGAT
TGACGGAATGTATTAACCTTACGCCCCACAAGTTGAGAGAATAATTGTTGAGATAATGGCTCAAATAAT
AGTGATTGATAAAAATATCAGTATTAAATAACCTGAAATTATTCTCGAAAACAAAGGCATTGC
ATTGCTCAGAACCATGGTGTAAAGAAGGGCCTGGAAGCAAAAGAGTTGACTATTATTCTCAGATC
AGGATACTGCTTCTAGCGATGTTATTGAAAACCTAACAGTACATTACGAAAAATAAAAGG
AAAATGTTGCTTGCTTCTCCTTTAAAGACCATCGTCAAATTATGCATCCGTAGTCAGCCT
AAATATTTCAGAGTACAAAGTTATGTTGAGTGAAGTAGACGATGATCTTATCCCTGCATGTTATTG
CTTCTGGGATGTTAATGTCCTGTAAGCAGTGGCGCGTGTGGACCATTGTTGAAAAACTCTTATAGAC
TGGGTTGATACAGAATGGTGTGCTGATTAGCTAAATAATGATTATTGTTGAGACACCATCAGTC
CATTTCTCATGAACTTGGGTATGGCAGAAAATTGCTGGTCGATCTGTTACAATACATAATTCTTCA
GAAATTTTATAAAATACGCAATGCAATATACTTAATGCTGATTCAAATTATAGCTCAAGTATGTT
CATGCTTTTTCATGCGACAAAGAATGTTGATTGAAATTATTCGAAAGAAAATTAAATTCACT
GAAGGTTGTTAAAGCTGACGTGATGGTATGTTCAATAATTGAAAGAAAATAGTTAGGCTCAAG
GTGTTAAATGGAAGAAAATAATGAAAGACGGTCGCTGAGTGGCACAGTGGGTTCTGTTGTT
GGTGGGTTGCAATCACTGTTGAGGTTATTGATTCAATCTGATGGTATACAATATCAGATATTG
CTCTTCAAAAATATGATAAAAATTAAATTAAATGCAAGGTTAATCTTGTGTTATTCAAAGG
CCAATGGCGTCTAGCATAATTGATATTGTTAAATTGTTATTCAAAGGCCAGATGTT
GTTTAATATTGGGGTGTGTTATTCTACCAATTATAAAACTATTCTAAATCAAAGGATTAT
TGTCAATATTGATGGGCTGTAATGGCGTAGAAAATAATGGGAACGTTGCTAAGAAAATTCTTAA
CTGAGGGGATATCTATTAGAATAGCTGATATTGTTATTGAGATAATCAAGCAATAGCTGATTGTT

AATAAGTACAAGAAAAAAGTAGTTATAGCTTATGGCGGAGATCATGCCACTAATCTTAGTACACCGAT
 AGACAATGATCAAAAAAAGAAGGTATTATTGGGGCTTGTAGGATAGAGCCTGAGAATAATAGAAA
 TGATTCTGAATGCCCTCATTAATACAGATAAAAAAAATTAAATTATGGGTAATTGGATAACAGCGAGTAT
 GGAGCCAGCTAAAAAAATTATCCTAAACTATCAAATATCACCCACTAGAACCTAACTATAATTGA
 AGAGCTTATAAAACTAAGAAAAATTGTCTGCATACATCAGAACACTCGGCTGGTGAACAAACCCCT
 CTTAGTTGAAGCGATGCATTTAATATTCTATTTGCTTCGATTGTGACTTTAATCGTTACACA
 AACAAATTAGCTCATTACTTAATGATTCTGAACAACCTAGCTTATTAGCAGAAAGTTGTCTTGAAA
 TCTTAAATGTCAGTATTAGATTAAAAATTATGCTGAAGATATGTATAACTGGAGGCATATAGCTGCTA
 TGTATGAATCTATTATTAAACGCATTAACAATAATATAATTGACCTTATATAGCAGGGAAAGATCACGA
 ACGCTGCCGCCGATCCCCATATGAATATCCTCCTAGTCCTATTCCGAAGTCCTATTCTTAG
 AGAATAGGAACCTCGGAATAGGAACTAAGGAGGATATTCAATGGATAAAGCGTAAGCATATAAGCATGG
 ATAAGCTATTACTTAAATAAGTACTTTGTACTTATTGCGAACATTCCAGGCCGAGCATTCA
 GCGGTGATCACACCTGACAGGAGTATGTAATGTCAGAACAGATCGGCTAGTCGTATGGCAGTGATG
 GGACGCAACCTGCGCTAACATCGAAAGCCGGTTATACCGTCTATTTCACCGTCTCAGGAGTTC
 GACGGAAGAAGTGATTGCGAAAATTCCAGGCAAGAAACTGGCTTACTATACGGTGAAGAGTTGTG
 AATCTCTGGAAACGCGCTCGCATCTGTTAATGGTAAAGCAGGTGAGGCACGGATGCTGATTGAT
 TCCCTCAAACCATACTCGATAAAGGAGACATCATCATTGATGGTGGTAAACACCTTCTCAGGACACTAT
 TCGTCGTAATCGTGAGCTTCAGCAGAGGGCTTAACCTCATCGGTAACCGGTGTTCTGGCGTGAAGAGG
 GGGCGCTGAAAGGTCTCTTATTATGCCTGGTGGCCAGAAAGAAGCCTATGAATTGGTAGCACC
 ACCAAAATGCCCGCGTAGCTGAAGACGGTAAACCATCGGTTACCTATATTGGTGCATGGCGCAGGTCA
 CTATGTGAAGATGGTCACAACGGTATTGAATACGGCATATGCAGCTGATTGCTGAAGCCTATTCTGC
 TTAAAGGTGGCCTGAACCTCACCAACGAAGAACCTGGCGCAGACCTTACCGAGTGAATAACGGTGA
 AGCAGTTACCTGATCGACATCACCAAAAGATATCTCACCAAAAGATGAAGACGGTAACACTGGTGA
 TGTGATCTGGATGAAGCGGCTAACAAAGGTACCGTAAATGGACCGCAGAGCGCGCTGGATCTGGCG
 AACCGCTGCGCTGATTACCGAGTCTGTGTTGACGTTATCTCTCTGAAAGATCAGCGTGTG
 GCATCTAAAGTCTCTGGTCCGCAAGCACAGCCAGCGACAAGGCTGAGTTCATGAAAAGTTG
 TCGTGCCTGTATCTGGCAAATCGTTCTACGCCAGGGCTCTCAGCTCGTGTGCTG
 AGTACAACGGATCTGAACACTACGGCAAATCGGAAGATTTCCTGCTGGCTGCATCATCGTGC
 TTCCCTGAGAAAATACCGATGCTTATGCCAAAATCCACAGATCGCTAACCTGTTGCTGGCTCC
 CAAGCAAATTGCCGATGACTACCAGCAGGCCGCTGCGTATGCTTATGCA
 CGGTCCGACCTCTCCGAGCGGTTGCCTATTACGACAGCTACCGTGTGCTGTTCTGC
 ATCCAGGCACAGCGTGAECTATTGGTGCCTACTTATAAGGTATCGATA
 AAGAAGGTGTTCCATACCGAATGGGATTAA

**SEQ ID NO: 13 (example O6A *rfb* locus nucleotide sequence – O6A-EPA production strain
 stGVXN4112 and stLMTB10923)**

ATGACGAATTAAAAGCAGTTATCCTGAGCGGGCTCGGGATGCATATGTTGCCACTAAGGCAT
 ACCCAAAGAGATGCTACCAATCGTCGACAAGCCAATGATTCACTGACATTGTTGACGAGATTGGCTGCAG
 GGATCAAAGAAATCCTCTGGTAACTCACCGTCCAAGAACGCCGTCAAAACCACTTCGACACCTCTTAT
 GAGTTAGAATCACTCCTTGAGCAGCGCTGAAGCGTCAACTGCTGGCGAAGTACAGTCCATTGCCGCC
 GGGCGTACAATTATGAACGTCGTCAGGGCGAACCTTAAAGGTTGGGCACTCCATTATGTGCACGAC
 CTGCCATTGGTACAATCCATTGTCGTTGCTGCCAGACGGTGTGATCGACGACGCCAGCGCAGCC
 CTGCGCTACAACCTTGCATGTTGCGCTTCAACGAAACGGCGCAGCCAGGTGCTGGCAAAACG
 TATGCCGGGTGACCTCTGAAATACTCTGTCATCCAGACAAAGAGCGCTGGACCGCGAAGGTAAAGTC
 GCCGCATTGTTGAAATCATGAAAACGGATCAGCGCAGACGCTGGACTCAGACATCATGCCGTTGG
 CGCTATGTGTTCTGCCGATATTGGCGGAACCTGAAACGCACTCAGCCTGGTGCATGGGGCGTATTCA
 GCTGACTGATGCCATTGCCGAACTGGCGAAAAAACAGTCGTTGATGCCATGCTGATGACCGCGACAGCT
 ACGACTGCGTAAAAAAATGGTTATATGCAAGCGTGTGAAGTATGGACTACGCAACCTCAAAGAAGGG
 GCGAAGTTCCGTAAGGGATTGAGAAGCTGTTAAGCGAATAATGAAAATCTGACCGGATGTAACGGTGT
 AAGAAAATTATAACGGCAGTGAAGATTAGCGGCAGAGTAATTGTTGCGAATTTCCTGCCGTTGTT
 TATAAAACAATCAGAATAACAACGACTTAGCAATAGGATTTCGTAAAGTTCCAGGATTTCCTGTT
 CCAGAGCGGATTGTAAGACAATTAGCATTGAAATTACGGGTTAGCGCAGTGTTAACGCTCGTAC
 ATCGTAGACATGCATGCAGTGTGCTGGTAGCTGTAAGCCAGGGCGGTAGCGTGTGAAATTATAAGTC

ATTCTTATAGAACATCGCATTCAATAATATAATTACACCTAAATGAATAGGATACAACGTGTGCACAATT
ATTAAAGGCTTAAAGATAAAAATAAAAAACGTATTTAGGGTTGTATATATTGCAGTTATTAATTATATC
GCGCATTGGTAAATTATCCTATCCTGATAAAATATATTGGGTTGGGGAAATGGGAATTGGTCTATAT
TACATCTATTATCAAATAGTGGCTTGATTATTGATTTGGCTTACTTACACAGGACCTGTGGTGTCTG
CGAGACATAGATGTGAGACCCAAAATTACAGCGCTTACTCAATAGTGTCTTTAAATCATTGCTT
TTTATAATTGCAATTACATGTGTATTTATTGTGCAGATTAAATATAGTCCACTTGTCTTTGGGTT
TTTGTCAATTTCATGCACATTGGTAATATATTATGCCAATTGGTTTGCAGGGGATTGGTGT
TTAAAAAACTTCATACTCACAAGTAATAGTGAGAATAACATTGTTATCATACTTGTCTTATGCTGT
AGTGGCGGAGATAATGTTTATCCTAAGTGTGCAAAATGCAACATTACTCATATGCTGTATACTT
ATGGCAAATATTCAATTAGCCATGGTGTCACTTAAACCTAATGAATGCATTGTGGAATTAAAGAAGG
CAGGAAATGTTTATTGGCGTAATAGGTACGATTGGTACAATGGCTAATTCTGTGTTAATTGAAAC
CTTGCGGTAATACGAGTCTGGTTTTCAATCGTCAAAAAATGACAACAGCATGCAAAGTCTAAT
TAATCCAATATCACAGTATATGTTATCTCAAGTTCAAGAAATTAAACCTCAAGATAAAACTGTTTATTATA
GAATTAAAAAGTTTTTGTGCATTAAACAATTAGCATAATTGCATGTTATGTTATATGGGTTAGGG
CAATATGTGGCGACTTTATAGGTAAAGTGCAGTTCAATTGTTATTATTTATTGCGTCAATAATTAC
CATTTCATTTAAATAATGCTTGGTATACAGTTCTTACCGACAGATAATGTAACAAACTAC
GAAGTATAATGTTATGGCGGAATTATTGTTAGTTGCTCTGGCTGTTAATATCAGCCTTGACATT
CTGGGGGGGTTTATTAAACCTAATTGGTAGTTCTTGTATTCAATTGCTAGCTTATTGCCATCG
AAAGTGGGGAGCGAGGTATAATGAAAGTGAAGGCGGTTCTGCTATTACATTCTATTAAAGTTAATGCT
GACAATTTCAGTGTACTGTTGGTATAGAACAAATAACAAATATCCTTGTATAGCAACGATAAA
CAGTTTTATATCGATATATCACTAATAAAATACTTCTCCGGCCAGCCTCTCGTTATATCATCTTT
GTGTTTATTGGTGTGCCCTTATTATCTTGTGCAAAACTATGATTATAGGATTGCCGATTGGTTAT
TGAAGGATATGGATGACGATGTGATTTGGCTAATGCTATAACACTAATGTTATGGTATACAT
TGGGACTAATTCTATGCAAAATACTGAAAAATTATCCGATGGCCTTACCTGTTAAGGAAACATTGCTA
AAAATAAGTTCTTGTACTTATTCTGGGTTGATAGGTATGGTGTAAAGGGATATTCTTTT
TAACTTATAGAACATCAATAGTTATGTTGATATTCAATCAAATATAACACGCCAATAGGTTATGATT
TTCTATCTTATTATTGTTCTTCTTCTTATGCGTTACAGTCAATTGCTAGCTAGTCAGAACAAATAAA
AAATTCTTTTATTGCGATATGCTTACAGTCAATTGCTGTTAAGGTTAGTCGTAGTGAAGCTATAAC
GTTCTTTACGGTACATGTATAATTAAATGAAGTAAAGACAAGAAACTACGCTGCTGATTACAA
TGATTCTTGTGTTAGCGTCATTGTTGATTAGTGAATTCTCAATGTGGCGCACTGGAGGGAGTTT
TTCAATTATGCAGGGTAATAATCCTGTTATAAAACTTGTATACGGCATGGAGTATCATATCTTCCAT
TTATCAATCAGTAAACACTACAATTGTCAGGGGATATAATGTTACCTATCTTCAAGCCAGTTAA
TAACCTGCTCGCAATTAAATGTCACATTGAGCTGCCGAAATAAGCTATAGCCATTGCCCTCATAC
ACAGCAACCCAGAACTATATAATTCTGGGTCGGACTGGGGGAGTTAGCAGAACGTTTAC
ATTGGTCTGATTGGATGTTCTTACCTTACCGCTTACTTAATTAAATGTTAGGAAACATTGCT
CAACAGATAATACAAATAGTGTATTAGCTGAACTGAAGGAAAAATATTCTCAGTTAGTATTAGAA
GAAACGAATAAAAACAGATGTATTCTAAACCGAGGGATTGAAATAGCCAAGGGAAATATTTTT
TGGTGTGATGACTCTTACCTCTTACCGGTGTTATCTCGGTTATTGGCTACAAATATGAGACAGGCG
CTGATGTAATCGCGCAAGAATACTTATATGAATAATAACGAGAAAACAATTGAAGATTGCATAATCGA
CATAAAAAGAGGGCGTTGTTAGTGTACAAATAGATTGGATTAGTTACATGTGATTGGACCA
TCCGATTGAATGTTTATGCACAGCCTTGTAGCTGAAAGGGAACATAATCGAAATATCGATTG
ATATATCTTACGGGAAACTGCTATCGTGAGGAAACTGATTCTGCTATCTTATTTAAATAAA
AAATTATATGATTCAAAGGCTTGTAAATAAAATTACCTCCAAGAAAAGCGACGGGAGGGCAAGAAC
AGCTAATCGATTAATCATTACGAAAGTGCATAAAATTATAGATTAAATTTAAAAAATATAATGATA
ATTGAAATCTCTTCAGGACAAAAGCATGCTATATTACCGACAGTGTCAATTGCTCTGCTAAAAATG
AAGTCGTTATCGGGAAAGTTTAAATGATTATATATCGCCGCGTATAATGGTTAGGGAGGGCAAGGT
GGGGTGGAAAGGGTTGTTGCCAACAATGTAACATTCTTAAATTTGGGGTTAAAGTCATTACTTGA
TAAAACATACTCAGGAAATTCTAACAAAATTGTAACAAAAAATACAAGTAGCAGTTATGACATGGCTATTGTTCT
CTTCTTATAGGAATGACATCTTAAAGCTCATGGCAATATGAAATGTTATTTCAAACAGTCATGAATAA
AAAACCTAATCGGTGTCGGCAGTGGCTTTATCTTCAAGCGTTGGGCTGGAGCATTTCAAAAAA

ATATCTGGCTGTTCAAATAAGGTAAAAGTGAATGGAATGAGCTTACAATATTAATTACACATAAAATC
AAAGTTGTCGAAATTATAAATCTGCACAATTGATTACACTGATGTTAATGAAGCAGAAATATGTGAC
ATTGTCGGCGATTGGAAAAAGGAAAGGAATAGATGATCTGATTACATATGAAAAATCTGCCAGATA
CTTCCTCCATTAGTTCAAGTATCCCAGCCCCACAAAATTGCTCGCTAAATAATGTTCTGACCAGC
ATTGCTGTCCCCATGCGAAAATGCCAGAAATATTAAGAAATCCAGAGTACTTATTACCGTCCTATT
TGAAGGATATGAGCTGGTACTATTGAAGCGCTATGCTGTGGTGCCTGTGATAGGCTATAATGTTGGTG
CAATTAGAGAGTTGATGAGAAAGTTCTGGCGTATTATTGCAATAATAAGAAGATTAGCACAA
GTAGCCTACAAATTAAATTAGTCTTGATAATGAAAATATTACATTGAGACAAACTATTATAGCAAGCG
TGAGCTTTCTGAAGAGAGATGCGGAAATTAAACGGCGCATTAATGAAAAAAATAAGAAACTC
TGTCTCATTCAATTAACTCATATAATGAACCTACCGGAGGGAGTATTTACGTACGCTGTTAGTT
TCTACAAAAACAGAATGTTAATTAAACACTATTGATAAAAATCCTCAGGTAAACTATTGAAAGACAATA
CTTTCAACATATATCATTTATTAAAGGTAACGTCAGGATATAATATCCAGGCTTTTTTATACCATCA
TTTATGTCCTTATATTCTCAATAATTAAAATTACGGAAGCAAGATATTCTGCTTTCAAACTC
TCGGCTGGATTGTTATGCTGCTTTAGAATACTCATGCCAACAAAAGATCATATTGTTACGGATA
ACTTCGAATATGACTTAATAAGACAAAAGATAAAACATAACTACTTTATTGAAAATTAAATTGTTAT
CTCAATGAATTATCAGGGCTTAAGAATTAGCTAGTTAGCTATATTACCCGGCAAGATAAAATGCAAT
GGATAAATTATGGATAAAAAAAAGCAGAAATTAAATTCTCCCTGTGATATTAGTAGAGAAAACCAA
CTGATGATTGTCAGCTACTTTATTAGTATAATGATTGAAATGATAATTAGGAAAAAGTAGTA
TTTACTGCATCTTGATTGTTCTGCTTTAGAATACTCATGCCAACAAAAGATCATATTGTTACGGATA
TAATGATTATTGCTATATTGAGGTAGGAAAGTACTACTTGAATCTCCTGATTGATAATT
TTTTTCGATAATCTATCAATTAGTGAATGTCAATTATTATCTGCTGTTGATTTTATTCTCCT
ATAGTTTAGGAAGTGGATGAAAACAAAATGCAAGACTATCATGGATTATATTTATGCGAC
AGAGCATTCTTAATCGGCTATGATGAAATTATACACAATAAGGAGTGTAAAGGAACTTCACATTGG
ATGAGGAATTCTAAAGATTCAAGATGAAAAGTATCAATAACAGCTAATAATGCTTACAGCAAAAA
TATTATTACATTATCGGTTAATGCCATGAACCTGATAATAAAATTGACGATTAGTTAGTGGAGAT
ATAATATGAACATATTAGTAACGGGGTGTGGATATATCGGATCTACGGCTATTGAAATTACTGAAT
GCAGGTCTGAGATTATCGTTCTGGACAATTCTAGTAATGCTCATACAAGTGTATGAAAAAAATAAAGA
AATTACTCGACGTGATTATAACAATTACTGGAGATGCTGGGTGAGGAAGACACTCTCGCTATTTCG
AGAAACACGCCATAGATAAGTTATTCATTGCTGGCTTAAATCTGTTAGCTAAAGTCTAAAGTGAACCC
TTAAAGTATTACCGAGATAATGTTGGAGTGACCTACTTATTACAGGTAATGGAAGAGTACAGAAATTAA
AAAATTATCTTAGTCATCTGCACAGTCTATGGTAACAGAGATAATTCCAATTCCAGAAACAGCTA
AAATTGGAGGAACCTACGAATCCATGGCACATCGAACGTTGAAAAATTCTAGAGGATGTTAGT
TCCACGGGAAACTGGATATAATTGCTTGAGATATTAACTCCTGCGGTGCTCATTCTAGTGGTAAAT
AGGTGAGGCTCCATCTGGTATCCCTAAATAATCTGTTCTTATTGATGTTGCGAGTGGTAAACGTG
ATAAATTATTATTATGGCAATGATTACCTACTAATGATGGAACAGGTGTAGGGATTATTATCATGTT
GTTGACTTAGCGAAAGGTCTTGGCTGCAATGAATTATTAAAGTATCAATTGGGATATAATATCTTAA
TCTGGTACAGGAAAGGTATTGGTACTTGAATTAATCACTACATTGAAAAATTAAACAAACATTAAAG
TCAATAATCTTTATAGAGAGAAGGGCAGGGGATGTTGCGTCTGTTGGCTGATGCAAGATAAGCTAAT
TCTTATTGGACTGGCAAGCGAACAAACTCTAGAACAGATGTTAGGACTCGTGGCTGGAAAAAA
TTATCCAGACGGATTCTGAATATAAAAGGTTCAGTTTATGAATCAATTGAGCAGAGAAAAAAACT
GGTTCTACACCTCGCTTCCCTACCCCTGTCTGGAGGGGATGATGAGCTATGTTATGAAAG
AACTTCCAAAAAAATGATCTTATTCTGAGCTATGTGATCAACCACTAGAACATTGAAATAAATATA
AATGACTCGGTCTCAAAGAAATTCACTGTCATCTACCAAAATAATCATATTATAATGTTAAG
AGCTTGGTACGCAAAACCGTGCACATTGCTTATTCAATGGAACACATTAAAGAATAAATACAATA
AATTAAATTAAACAATGCGATGCACTATTGTCATCTGATAAGAGTGTGATTGTTAAGGATACAGAC
AAGTCAAAATTCTGATATGACAGATGCAATATCTTGAATTACAGTCGCGTAAAAAATTAGCAAGTAA
AAAAAGTTGCGTGCATTATTATCTGGAAACAAAAAGATTAGAATCATATGAACGTTCTGGCGA
ATCTTTGATTGACCACTTTATTCTCGTAGACCGTGAATCTCTACCCATACTGGCAGTAAT
ATCCATATAGTCATAATGGGTTGATACATCAGCCTGAGATATAAAAAGAGAAATAAAATCGATAA
GCCTGTGGAACCTATATTATCGGAAATATGATTCTTACAAATATGGATGCTGCAAACATTTGCTA
AGAATATTTCACCTGCTGTGATGAGTTAATTATTAAAGTGTGTTAAGGATCTCAGAAACT
AATAAAATATTAATTCAATTAAAATAACATTGCTTAGGTTAGGACTGTTGATGATCAATTCTCCGC
TTCTACAGGGCATATAGGTATATGCTCTGTTGAGCAGGCGTACAAAATAAATTCTGAATACA
TGGCTTAGGTTACCATGTTACATCTGAGTTAAGGATTAATGCAAATCAGGTAGCGAA
ATTTTGTTGCGAGATACTAGAGCAATAAAAACGTACTAAGAGAAATAATTACGATTATAATCGTTA

TAATGAATACATTAGATGAGAAATTATATGAACAATAATAAAATTATTACACCTATCATTATGGCTGGTGG
TTCAAGGCAGTCGGTTGTGCCACTATCAAGAATTCTCATCCGAAACAAATTCTTAGCCTAATCGGTAGTC
ATACCATGCTTCAAACAAACGGCTAACGCTAACTCGTCTGGATGGTTGGATTGTACCAACCCTATGTCATTGTAAT
GAACAATACCGCTTATAGTGCTGAACAGCTAGAAAAATCGATAGATTGACTCAAAGAATATCATCCT
TGAGCCTGTTGGCGTAACACTGCCCTGCAATTGCATTAGCGCGTTGCTGATGTCTAAGTCTGATAAAA
GTGCAGATGATCTTATGCTCGTACTGGCTGCAGATCACGTTACACGATGAAGAAAATTGTAACATTGGTATAATTCCAGACAAAGCAGA
AACTGGTTATGTTATATACATCGAGGACAATATATTAACTCAGGAAGATTGGATGCATTATAGTGT
CATTGTTGAAAGCCAATCATGAGACAGCCACTAAATATCTGCTTCCGGTGAGTATTATTGAATAGC
GGTATGTTTGTGAAATCGTTATAGAGGAACCTAAACAATTCCGGCTGATATTATCCGC
TTGTGAAAAGCAATTGCTTCAGCGAACCTTGACCTTGATTTGTGCGTTAGATGAAAGTCTTCTCTA
AGTGCCCTGAAGAATCAATTGATTACGCTGTAATGGAAAAACAAAAGACGCAATTGTTATTCCAATGGAT
GCTGGCTGGAGTGATGTCGGTTCATGGTCTCTTGGAAATTATGATAAAGACTCAGACGGCAACGT
AATAGTTGGGATATTTCTCTCATGAAACAAAGAATTCTTCATATATGCCAATGGGATTGTTGCTA
CAGTGGAGTGAAAATTAGTTGCTCAAAACAAAGGATGCTGTTCTCAGAGAGAAAATAAGTT
CAGGATGTAAGAAAATAGTAAACAAATTAAAATTAGGTCGTAACGAGCATTGTTCATCGCGAAGT
ATATCGTCCTGGGTAATATGATCCATTGACACAGGGGAGCGTATCAGGTCAAACGTATAACAGTAA
ATCCTGGTGAAGGACTTTCTTACAAATGCACCATCATAGGGCAGAACATTGGATCATAGTTCTGGAACT
GCAAGGGTGAACTATAGTTCTGAAACTAAGATTCTAGCGAAAATGAATCTGTTACATACCTTGGTGT
AATACACTGCTGGAAAATCCAGGGAAATTCTCTTGTGATTTAATTGAGTTGTTCTGGATCTTATTAG
AAGAAGACGATGTTATCCGTTTCAGGACGATATGGTCGTAGCTAAATTGATAATGTAACGTTAGTA
GAAGAGCGCTAATATTTAGTTAATCTGTAATAAGTATTATTGTTAAGGTATATCATGTCGAGTTAC
CCTGTTAAAGCCTATGATATTGCGGGAAATTAGGCGAGAACTGAATGAAGATATTGCCCTGGCGCATT
GGTCGCGCTTATGGCGAATTCTCAAACCGAAAACCATTGTTAGGCGGTGACGTCCGACTCACCAGCGA
AACCTAAACTGGCGTGGCGAAGGGGTTACAGGATGGCGCGTCGATGTGCTGGATATTGGCATGTCG
GCACCGAAGAGATCTATTGCCACGTTCCATCGCGTGGATGGCGCATCGAAGTTACGCCAGCCAT
AACCCGATGGATTACAACGGCATGAAACTGGTGCAGAAGGGGCTGCCGATCAGCGGTGATACCGGACT
GCGCGACATCCAGCGTGGCAGAACGACTTCCCTCCCCTGATGAAACCAAACCGGGTCCGCTATC
AGCAAATCAATCTGCGTACGCTTACGTTGATCACCTGTTGGTTATATCAACGTAAAAACCTCACGCC
CTCAAGCTGGTATTAACTCCGGAACGGCGCGGGTCCGGTGGACGCCATTGAAAGCCCCTTAA
AGCCCTGGCGACCCGTTGAAATTCAAAGTCACACGCCGGACGGCAATTCCCCAACGGTATT
CTAACCCGCTACTGCCGAATGTCGCGACGACACCGCAATCGGTCATCAAACACGGCGGGATATGGC
ATTGCCATTGATGGCGATTGACCGCTGTTCTGTTGACAAAAAGGGCAGTTATTGAGGGCTACTA
CATTGTCGGCCTGCTGGCAGAACGCTTCTCGAAAAAAATCCGGCGCAAGATCATCCACGATCCAGTC
TCTCTGGAACACCGTTGATGTTGACTGCCGAGGCAGGCCCCGTAATGTCGAAACCGGACACGCC
TTTATTAAAGAACGTATGCGCAAGGAAGACGCTATCTACGGTGGCGAAATGAGCGCCCACCAATT
TGATTTCGCTACTGCGACAGCGGCATGATCCCGTGGCTGCTGGTGCCTGAAAGGAA
AAACGCTGGCGAACTGGTGCAGCAGCGGATGGCAGCGTTCCGGCAAGCGGTGAGATCAACAGCAA
GCACACCCGTTGAGGCATTAAACCGCTGGAACAGCACCTTACGCGTGGCTGGGATCGCAC
CGATGGCATCAGCATGACCTTGCGACTGGCGTTAACCTGCGCTCTAACCGAACCGGGTGGTGC
GGTTGAATGTGGAATCGCGCGCGATGTACCGCTGATGAAAGAAAAGACAAACTTATCCTGAGTTACTG
AACAAAGTAATTCACTATATAATGGTTTAAAAACGGAAAAGATGAGATATCCGGTGTGGTAT
ATCCAAGGTAATGCTATTCACTGAGTTAACATCTATACCACTTAAGCCGACACTTC
GGGATCCCCATATGAATATCCTCTTAGTTCTATTCCGAAGTCCCTATTCTTAGAGAAATAGGAACCT
CGGAATAGGAACTAAGGAGGATATTCAATGGATAAAGCCGTAAGCATAAGCATGGATAAGCTATT
ACTTTAATAAGTACTTGTATACTTATTGCGAACATTCCAGGCCGAGCATTAGCGCGGTGATCACAC
CTGACAGGAGTATGTAATGTCGAACAGAACGATCGGCGTAGCGGTATGGCAGTGATGGACGCAACCTG
CGCTCAACATCGAAAGCCGTGGTTAACCGCTCTATTTCACCGTCCCGTGGAGAACGCGAACAGT
ATTGCCGAAAATCCAGGCAAGAAAATGGCTTACTATACGGTGGAAAGAGTTGTCGAATCTGGAAAC
GCCCTCGCATCCTGTTATGGTGGAAAGCAGGTGAGGCCACGGATGCTGCTATTGATTCCCTAA
ATCTCGATAAAGGAGACATCATGATGGTGGTAACACCTTCTCAGGACACTATTGTCGTAATCGT
GAGCTTCACTGAGAGGGCTTAACTCATCGGTACCGGTGTTCTGGCGGTGAAGAGGGGGCGCTGAAAGG
TCCTCTATTATGCTGGGCCAGAAAGAACCTATGAATTGGTAGCACCAGCTGACCAAAATGCCG
CCGTAGCTGAAGACGGTGAACCAGCGTTACCTATATTGGTGCGCATGGCGCAGGTCACTATGTGAAGATG

GTTCACAAACGGTATTGAATACGGCGATATGCAGCTGATTGCTGAAGCCTATTCTCTGCTTAAAGGTGGCCT
 GAACCTCACCAACGAAGAACTGGCGCAGACCTTACCGAGTGGAAATAACGGTAACCTGAGCAGTTACCTGA
 TCGACATCACCAAAGATATCTTCACCAAAAAAGATGAAGACGGTAACCTGAGTGTGATCCTGGAT
 GAAGCGGCTAACAAAGGTACCGGTAATGGACCAGGCCAGAGCGCGCTGGATCTCGCGAACCGCTGCGCT
 GATTACCGAGTCTGTGTTGCACGTATATCTCTCTGAAAGATCAGCGTGTGCCGCATCTAAAGTTC
 TCTCTGGTCCGCAAGCACAGCCAGCGACAAGGCTGAGTCATCGAAAAGATTCGTCGCTGCCGCTGTAT
 CTGGCAAAATCGTTCTACGCCAGGGCTCTCAGCTGCGTGTGCCGATCATCCGTGCCGAGTTCCCTGCAGAAAA
 TCTGAACACTACGGCAGAACATCGCAAGATTTCGCTGCTGCCGATCATCCGTGCCGAGTTCCCTGCAGAAAA
 TCACCGATGTTATGCCGAAATCCACAGATCGCTAACCTGTTGCTGCCGACTTCAAGCAAATTGCC
 GATGACTACCAGCAGGCCTGCGTGATGCGTTATGCAGTACAGAACGGTATTCCGGTCCGACCTT
 CTCCGAGCGGGTGCCTATTACGACAGTACCGTGCTGCTGCCGACCTGATCCAGGACAGC
 GTGACTATTGGTGCCTACTTAAAGCGTATCGATAAAGAAGGTGTTCCATACCGAATGGCTGGAT
 TAA

SEQ ID NO: 14 (example O8 *rfb* locus nucleotide sequence – O8-EPA production strain
stLMB11734)

ATGACGAATTAAAAGCAGTTATTCTGTAGCGGGCTCGGGATGCATATGTTGCCCTGCCACTAAGGCAGAT
 ACCCAAAGAGATGCTACCAATCGTCGACAAGCCAATGATTCACTGAGTACATTGTTGACGAGATTGTTGGCTGCGAG
 GGATCAAAGAAATCCTCTGGTAACTCACCGCTCCAAGAACCGGGTCGAAAACCAACTTCGACACCTCTTAT
 GAGTTAGAATCACTCCTTGAGCAGCGCTGAAGCGTCAACTGCTGGCGGAAGTACAGTCCATCTGTCGCC
 GGGCGTGAACATTATGAACGTCGCTCAGGGCAACCTTAAAGTTAGGTTAGGCCACTCCATTGTTGCGCAG
 CTGCCATTGGTACAACCCATTGCTGCTGGTACTGCCAGACGTTGATCGACGATGCCAGCGCCGACCCG
 CTACGTTACAACCTTGCTGCCATGATTGACGTTCAACGAAACGGCCGAGCCAGGTGCTGGCAAAACG
 TATGCCGGGTGACCTCTGAATACCTCCGTCATCCAGACTAAAGAGCCGCTGGACCGTGAGGGTAAAGTCA
 GCCGCATTGTTGAATTATCGAAAACCGGATCACCGCAGACGCTGGACTCAGACATCATGGCGTAGGT
 CGCTATGTGTTCTGCCGATATTGGCCGGAACCTGGAACGTAACGACTCAGCCTGGTGCATGGGACGTATTCA
 GCTGACTGATGCTATTGCCGAGCTGGCAGAAAACAATCCGTTGATGCAATGCTGATGACCGGCACAGTT
 ACGACTCGGCAAAACGGTATATGCAGCGTTGAGTATGCCCTACGCAACCTGAAAGAACAGGG
 GCGAAGTCCGAAAGGTATTGAGAAGCTGTTAAGCGAATAATGAAAATCTGACCGGATGTAACGGTTGAT
 AAGAAAATTATAACGGCAGTGAACCGGATTCGAGCAAAAGTAATTGTTGCGAATCTTCTGCCGTTGTTTA
 TATAAACCATCAGAATAACAACGAGTTAGCAGTAGGGTTTATTCAAAGTTCCAGGATTTCTTGTGTT
 CCAGAGCGGATTGGTAAGACAATTAGCGTTGAATTTCGGGTTAGCGCAGTGGTAACGCTCGTCAC
 ATCATAGGCATGCATGCACTGCTGGTAGCTGTAAGCCAGGGCGTAGCGTGCATTAAACCTCTATT
 AATCAAACGAGAGCCGTTATTACAGCATGCTCTGAAGTAATATGGAATAAAAGTAGCTAGCGATCGC
 TTAAGATCTAGGATTTCATTATGTTACTTCCTGTAATTATGGCTGGTGTACCCGAGTCGTCCTCGGCC
 ATGTCACCGAGCTTATCCGAAACAGTCCCTCCGCCTTCGGGAGAACACTCCATGCTGCAAGGAAACCAT
 CACCCGACTCTGGGCCTTGAAATCCATGAACCGATGGTCATCTGTAACGAAGAGCACCGCTCTGGTGG
 CTGAACAGCTACGCCAGCTCAATAAGCTGCGAATAATTATTCTGAGCCGGTGGCGAACACCGCC
 CCGGCCATGCCCTGGCAGCCCTCAGGCCACCGCGACGGCAGACCCGCTGATGCTGGTCTGCC
 TGACCATATCATCAATAACCGAGTCGCCCTCCACGACGCCATCCGGTCGCCAGCAGTATGCTGATGAAG
 GTCATCTGGTACCTCGGTATCGTGCAGTGGCCGAAACTGGCTACGGTTACATTAGCGCAGCGTGTG
 GCGCTCACCGATAGTGCCATTCCCGTACAGGTGGCCGCTTGTGGAGAACGGGATCGCAGCGC
 CGAGGCTTACCTCGCTCCGGGAGTACTACTGGAACAGCGGATGTTATGTCGCCAGGCTGCGGTGAATGCC
 TCATCGAGCTGCCAAATACCGTCCGGATATCTGGAAAGCCTGCCAGGCTGCGGTGAATGCC
 GGCAGCGATTTCATCAATAATCCCGCATGATATTCTGCGAGTGGCCGATGAGTCCGTGGACTATGCC
 TATGGAGAAAACCGCGATGCGGTGGTGGTCTCGATGCTGACTGGAGCGACGTCGGCTCTGGT
 CACTATGGGAGGTCAAGCCGAAAGACGAGCAGGGCAATGTCCTCAGCGGTGACCGTGGGTACACA
 GAAAATGCTACATCAACAGCGACGAGAAGCTAGTGGCGCCATTGGCGTAGAGAAATCTGGTATTGTC
 CACTAAGGACGCCGTGCTGGTGTGAATCGCGAGCGTCCAGGACGTGAAGAACGGGTCAGTT
 AGCAGAACCGCGCAGCGAGTACAAGGCCACCGTGAGATTACCGCCCTGGGCCGTTGCGACGTAGTG
 GTCCAGACCCCGCCTCAACGTCAACCGCATCACGGTGAACCGGCCACCGGTAGGTGACTGCA
 CCACCATCGCGCCGAGCATTGGTTATTCTGCCGGCACCGGTAGGTGACTGCAACGGTAAGCAGT
 TTGTTGCCGAGAACCGAGTCCACCTTATTCCGATTGGCGCCGAGCAGTGCCTGGAAAACCTGGCTGTATT

CCGCTGGAAGTGTGGAGATCCAGTCGGGGCGTACCTGGCGAGGACGACATTATCGTATTAAGACCA
GTATGGTCGTTGCTAATTATTCGGGACAAGACGAGAATGACACAGTTAACGGTTAAAGCTTATGA
CATCGTGGTAACTGGGTAGGAACACTGAACGAGGACATCGCCTACCGTATCGGCGCCTACGGCGAAT
TTCTGAAACCCGGGAAGATACTGGTGGGGCGATGTGCGCCTCACAGCGAGTCGCTGAAGCTGGCGCTG
GCCCGGGTTAATGGACGCCGTACCGACGTGCTGGACATCGGCGAGCCACAATCTATGAACACTAACG
GCATGAAGCTGGTGCAGCGAGAATCGAAGGCCATCAGCGCGACACCGGCGTGCAGGAGATCCAGCGCCTG
GCGGAGGAAAACCAGTCCCGCCAGTGGACCCGGCGTGCAGGAGCCACTGAGCAAGATACTGGTACTGAA
GGAGTATGTTGACCATCTGATGAGCTACGTGGACTTCTCGAACCTCAGCGTCCACTGAAGTTGGTGGTGA
ACTCCGAAACGGGCTGCCGGCACGTATTGATGAGGTGGAGAAACGCTTCGCCGGCTGGGTGCCG
GTAACCTTATCAAGGTGCATCACCAGCCGGATGGCATTCCCTAACGGTATCCCAGTCCGCTGCG
GGAGTGCCGCCAGGATACCGCCGACGCCGGTGCAGCGAGCATCAGGCCAGATGGGATTGCCTTGACGGCG
ACTTCGATCGCTGCTTCTGTTGATGACGAAGCTTCGTTATCGAGGGGTTAACATTGTCGGCCTGCTG
GCTGAGGCGTCCCTGCAGAACGAGCCGGAGCGAAAATCATTACGACCCGCGCTTGACGTGAAACACGGT
AGACATCGTACCCGCAACGGCGCAGCGGTGATGTGAAAGACGGGAGTCGTTACAGTCAAGGAGCGGA
TGCAGGAAAGACGCTATCTACGGGGAGATGAGTGCACCAATTCTCCGCGATTTCGCTACTGC
GATAGCGGGATGATCCCGTGGCTGGTGGGGAGCTGCTGTGTAAGAACAGCTCGTGAATCGCT
GGTGGCGGACCCAGAAGGCGTCCCTGCCTGGAGAGATCAACCGCAAGCTAAGTAATGCTGCTGAGG
CGATCGCCCGCATCCGGCGCAGTATGAGCCGGCGCTGCACACATCGACACAACGGACGGGATCAGTATT
GAATACCCCTGAATGGCGCTTAAACCTGCGCACGTCAACACCGAGCCGGTGGTGCCTGAACGTTGAGTC
CAGAGCTGATGTGGCGTTATGAATAAAAACGACCGAGCTTACACCTGTTAACGGGAAATAAGGTG
AGAGATTTACTAACGACGATTTATCGTTATCGGGATTATCTGGAGCAGTGTAAACGTGATTTCAAGG
ACGCTATCAAACACTAGTATGCTGGCGACTATGGCTCGTTAACACCCTCTATGATTCTGGTCTATA
CCCTGGTTTTCCGAGGTGATGAAGGCAAGAATGCCGATAAACCGGGTCTGGCTATAGTATT
CTCTGTTCCGGGTTACTGACCTGGGATTATTACTGAGATGCTGGATAAACGGTCAAGGCTATT
CAATGCTAATCTGATCAAGAAACTCAGTTTCCGAAAATCTGCTGCCGATCATGTCACGTTATCGCG
TGCTAAATTGCGATTATTCAGTCTGTTCTAACCTTATCATGTCACCGTAACCTCCCGGCTGG
CTCTTCTCTCGGTGATACCGGTCTGCTTTGAGATCTGTTGCCGGTGGGCTGGGATGATCCTGG
TGTGATGACGCTTTTCAGGGATGTGGGGCAACTGGTTGGCGTGCCTGCAATTCTGGTTGGTCA
CACCCATTGTTATGACTGAATTCTACCTGCACTGGGAAACTCTGATGAGTATAACCCGATGACT
CGGATCATGCAATCTTACAGTCCATCTCGCCTATCATCTGGCCCCAAGTGGTATTGCTATGCCAGT
ATTGGCTCTGCCATTATTCGCTCATCGTTCAGGATGTTCCGCAAGCATGCCGGATATGGTGG
ATGAATTATAATGAGTTATCAGAGTAATAATGCGGTAAAGCGTATGCCAGTACTCAAAAGACCG
GGAGACTGATGAAATGGTATCCCTCTGAATACCAACGCCATAATTGAAATGGATCCTCGCGATATT
AATTGCAAGTCGCTCCGGCGAGGGTGTGGTATTATCGGTATCAACGGTGCAGGCAAGAGTACCTGCT
TAAACTCATAACCGGGACGCCAGGCGACTGGAGAAAATTGAAATCTCCGGACGTGCGCTGCATTAC
TCGAATTGGGATGGGTTCTGATTCACTGGTCCGAGAATGTTATATGTCGGCAACTGTTG
GGGTTATCGTCAGAGAAAATAACTGAACTGATGCCGAAATTGAAAGAGTTGCTGAGATTGGGACTATAT
CGATCAACCTGTCGCGCTACTCCAGTGGATGCAAGTTCGATTAGCTTGTAGCGACGGCTATCC
GTCCTGATGTGTAATTATCGATGAGGCATTATCTGTTGGGATGCAATTTCAGCATAAAAGCTTGG
CGTATTGCAAAATTGCTCAGGAAGGGACCACGCTGTTGCTGGTATCCATGATAAACAGCGATCCAAG
CATTGCAACGGGCCATTATTAAGGCAAAATTGAAATGGAAGGTGAACCTGAGCAGTGTGAG
ATTATTACAATGCTCTCTGGCGATAAACAAAATCAGTCCATTAAACAAAGTGTGAGCATAATGGTAAACG
CAAACCTGTTCAAGGCACTGGTGGACTATCTGAGGTTACGTTCTCGATGAAACAGGGCAATGTGAC
TGAATTGTTCCGGTAGGGCATCGTGTCACTGGCAGGTCACGTTGAGGTCAAGGACGATATTCTGAGC
TTGTTGCGGATATGATTAAGGATCGACTGGCAGGGGAGGACATTTCGGGACCAATACGTACCATCTCAAT
CAGACACTCACCTCCCTGAAAAAAGGAGAAAAGCGTTGCTTCTTATTCGATGCGAGATTGGGGT
TGGCTCTATTCTGTCGCTGCGCTTGCATACTCCAGTACGCACCTCGGAAAAACTATGAATGGCG
ATCTGGCGTGGTATTCAACGCTGTTAACACCGAACACAAGAGTTGTCGGCGTGTCTGGTTGCCG
GAACTGGAGATTCTTAATGGGTCGTTGCTGTTATCGTCATTGAAAGAACGACACAGAGGTCGGTGA
GAAATCAAGGCCGCTTGAGTTTATTACCTTCTGCAAGGTCGAAGGACATTTACCTGATGGCGT
GATTGCGGATATTGGTTGCGGACGTGGCGAATGGTTGGAGATCCTGACTGAAATGGCATTGCAACATCG
GCGTCGATCTCGATGATGGCATGCTGGCGCGCCAGGGAGGCGGACTGAATGTGCAAGAAAATGGATTGT
CTGCAGTTTGCAAAGTCAGGCGGATCAGAGCCTGATAGCGTTGACCGGTTTCATATTGCTGAGCATT
GCCGTTGAGGTCTGCAGCAACTGCCATGCATACCCCTACGGGTGCTGAAACCAGGTGGTTGCTGATCC

TCGAAACGCCGAACCCGGAGAATGTAAGCGTCGGCACCTGTCATTATGGATCCAACGCATAATCAT
CCTCTGCCACGCCACTGTTGAGTTTACCTATTCAATTGGTTTACCGAGCAATTACCGTCGTCT
GCAGGAAAAAGAGGTCTCAATCTCGGATGCAGCGTAATTGGTCATGACTCAAAGGGTGAGCC
CCGACTACAGCATCATTGCTCAGAACAGCAGGCCAACAGATATTCTGAACGCTTGACACCCCTGTTACC
CAGCAGTACGGTCTGACGCTGGATGCTCTGAGCAACCGTACGATGCGATTGCGCAACAGTTTCGTC
CGTTGTCACGGCTGGAGACGTTGAAACCAAACCTATATGCAACAGATAAGCAAATGTCAGAGACTATT
AGACGTTGCAAGGTGAGGTGACGATCTGAGTCATGATCGATCAGAACCATCAGCTCATCAGCAAATG
GCGGATTTACATAACAGTCGTTCATGGCGTATTACTCAACCAACTACGCTGGTGTCTTGCAACGTCAATT
ATTACGTCAGGAAGGGCTAAAGTGCAGGCCGTAGGGCTGGGAAAAAAATTGCGCAAAGGGATGGCGC
TCTCGCTGGTCTTTCCATCGTACCTAAGGTTATCTGTTAAGGTTCTGAGAAAAACTGGC
TGCTATACATTGCTACAACGTTGTCACCGTAATGCTGGTCAATCTGACACGATGATGATGAGTC
CAGAAGATATGATGTTGGTACTGAAGAAATGACAAGTCGCGCATGAGTATTATAACGAAATTAAAATA
AAAATACGGAGAAAATAACGATGCGTATTGTCATAGATTACAAGGCGCACAGACGGAAGGCCGTTTCGTC
GCATCGGTCTTATAGTATCGCAATGCCAGAGGCATAATCAGAAATAACAGCCGCATGAGATTTCATC
GCGCTATCCGCATGCTGGATGAGTCGATTGCAAATATTAGCGCAATTGCGATCTCTGCCGGCAGA
AAATATAGTCGTATGGCATGCCGTAGGCCCTGTCGTGCGATGGACCAAGGTAATGAATGGCGTGGGAGA
GCGCAGAACTGATTGGGAAGCGTTCTGAACTTACAGGCGATGTCAGTAAAGTAGCCGTACTGCA
CCACGATCTTATCCCCCTCGTCAGCGGAAACCTATCTCAGGACGATGATACAAACCTACTATTAC
AGAAAGTTGAGTGGTAAAAAACGCTGACCTTTGTTGACTAACTCTGCTTATACCGCACAGGAAGCGATC
GAGCATCTGCATTACAGGGCGATCATGTCAGAATATTGCAAGCCGCAGTCGATTCTCAGTTGTATGGC
GGAGGTGGCAGCGAGCGAAAAGAGACCGTCTGGCATTACGGTATTAGCAGCGCAGTTCATGTTGATG
CGCCCGAGGATTGACTCAAGGAAAAACTTAAACGGTGTGATTGAGGCCTATGCCGGCTCAGTGATGCC
TTACGTCGAGTCATCAACTGGTCATCGTCAGTAAGCTTCCATCGGTGATCGTCAGTATCTGGAATCCCT
TGCAGGTAATGGTTACAGCAGGGCGAAGTGGTACTCACTGGTTATGTGCCGGAAAGATGAGCTGATCC
AGCTCTATCGCTATGTAAGCTGTTCATTTGCTTCACTACATGAAGGTTGGGTTGCCGGTCTGGAA
GCAATGTCGTGCGGTGCGCCGGTATTGGCTCAAATGTCACCGATATTCTGAAAGTCATCGTAATCTGA
GGCATTATTGACCCGTATTCTGTCCTCCATGAGGGATAAGATCGCGCAATGTTGACTGATGATACCT
TCCTCGCGCTGAAAGAAATGGCGAGCAGCAAGCGCTAATTCTCTGGATAAAAGCTGCCGGTACT
GCTCTGGAAGCTTCGAAAGATCGCGTAGAAGACACCGGTACTGCGCAGGTTTGCGCTGAAGCTTGT
TCAGAAGATCCTGCTATCTCACAAGGGCAGCCAGATGACCGCGATCTGCGCTGTGCGCAACGCCATTG
ATTACAATCTGAAACGGCAGAACTTATCAAATGACGATAATCGCTGAACGGCGTGTGGAAGGCCA
TTCGATAGCTCATATAGCTGGCGTGGTCAACCGCAATTGCCCCGGCACTCTCAGCCGATGGTGTAGA
GGTTTATTGCACTTCACTGAAGGACCAAGGTGATTGCCCCAGATGCCGTTATGGCACAGTCGGAAA
ATAGTGATCTCTGGCATTATAATCAATGTCAGACCCGAAGAGTAACGAAAAGATAGATATTAGC
AGAAATATCTACCCCGGGTTACCAAAATGGATGCCAAAGTAAAATTCTCATTGTTATGCTTGGGAA
AGAAACGGGTTCCGCAACCGTGGATCAATGAATTAAATCGGAACCTGACGGAGTGCTGTACTCGG
AACATGTCGAAACTGATTGATAACGGACTGAATGTGCCGCATTGTTGGCAATGGCTGTGAC
CATTGGCTCAATATCCCAGCCGAGACGACAAAAGATGTGGATCACGGAACATTCCGTTCTGCACGTCTC
TTCTGTTCCACGCAAAGGGATACAGGCAATGCTTCAGGCTTGGGGAAAGCGTTCACTCGTGTGACA
ATGTTATCTTAATCATTAAAGACTTTAACAAATCCGACAAATGAAATTGACGCATGGCTGGCTCAGGCCAG
GCTCAATTGACTATCCAAAGGTGAAGTGATCAAAGAGGATATGTCAGCCACCGAGCTAAAGGGCT
TTATGAAAGCTGTGATGTTGGTGTCCAGGTTGCGCTGAAGGCTTGGTTACCTATTGCTGAAGCAA
TGCTGAGTGGCTACCGCTATCGTCACCAATTGGAGCGGGCAACTGATTGTTGTAATTCAAACAAATTCA
TGGCTGGTTGACTATCAGTCAGCTGGTAAAGCAGCACTTGGTCTGTTCTCAGCCTGGCCAGTGT
GGATATTGACAACAGATGCAATTAAAGCGCAGCCTCAACCGATAAATCAGTGCTGCGTACATGG
CCAATGCTGGTCGCGAGCTTCTGCAAGCAGTTACCTGGAAAGCGGTGGCTGATGTTCTGCCAGCG
GTCAAGACTCTCGTGCACGATATTGATATTGACACAGCAGTCGGCGCATGGCTGGGTGACGACCTGGAA
CACGAAATGTGGATCGCAACCTATCCCAGCATCTGGTGGAAAGCGCACCTCATGGCGGGATGTTGTT
TTGCTCCCCAGGTAGCGCTGGCGATCTGTTGTCAGACGAAGAGTTGACTTCGCAACTGGATTGTA
GGTAAAGAGAGCAACTATGGAAAACCTCCAGCCACACATTGATGCTCTGAGACTCGATGTCATTGAT
CCAATTCAACTATGGATTCTTAATCATGAGAACTGTCGGCTTATTGTCGCCAGCATGACGCCGTC
GTTCAAGTGTGTTATGACGATGCACTCAACTGTGGATCCGCTGGAAAAGAGCCGAGCTGGAATTCCGCTT
GCTGAAATGAAAGAGCGCTGGCACCTTGCAGCCGGTTGGTGCATTGCGATATGAACCGCCT
TAAAGATTAGGCTTAACGCAATTGCTTATTCCCGCACGGTGTATCAACTACTCCGAGCGAGCG

TCACACGTCAACAGCAGTCTTACCGCTAATTGCGAGCTATGGCTTCTGCTTACCGCATAAGGGCCTGATG
GAACTAGTAACTCCGTCCATAGACTCAAGCAAGCCGGTAAACCGGTTCTGTTACGACTGGTGAACGCAGA
GTATCCTGTTGGGGAGTCACCGCATCTGGTGGCAGAGCTAAAGCTGCTCAGCGGTAGGTGTTACCG
ATCTGATTGAGATGCATAATGATTCCCTACCTGATGCCAGAGCTGCGGTTGCTTACAGAGCCGATCTT
CTGATTTGCTTATCAGAATACTGGGAGTCAGCTAGCGGGCGGTACGTTATGGTATGGCAGTCAAAA
ACCTGTTGCGGTAAACGCCCTGGCGATATTGATGATTGGACGATGCCGTCTTAAATTGATGGATGCA
GCGTCGATGATATCAGTCAGGGGATTGACCGGATCTGAATTCCATCCGTAAACAGAACTCTGGCAACC
AGGAACTCAACAACGTGCCATGCACTGGCGGGAAACAACATGATTATCAAGCTGTTCACGCCGTCTGGTAA
TATGTGTCAAGGCTTAGCTAAAGCTAAATATTAAATAAAATATCTCTCTGTATTTTGCTACTGAGCCAATTAAACCA
TACAAGAGGGGTTAGATAATGTGTCAATTATTGAAATTATTTGCTACTGAGCCAATTAAACCA
TTAACGGGCATGGTCGGTATTCCCTGGAGCTGGTTAACGGGCTGGCGGCGAAATTGAAGAATT
AAAGCTATTCACGGTGCCTGTTATAGAACAGATCCCTTGGTGGAGAATAAAAGCGATACCAAAGCCA
GCAATCATGGTCGTCGTGCGCGTTCTACGCCGACAGACGCTGTTGATTGAGGCTTATCGCTGCTGCAT
CCGCGCGCCAGGCCTGGCATTGCGCAGTAAAGGATTATATCTACCATGGCCCCAATTATCTGCC
GCATAAAACTGGAACCGCGCGTACCGACGTTCATGACATATCCATTTCACCTGCCGGAAATCATCCAA
AAGATCGGGTCGCTATATGGAGAACGATCCCTGCATGAGAGCTGGATTGGCAAAGCTGATCCTGACCGTT
TCTGATTTCGCGCAGTGAATTATCCGCTTGTCAACTATCCGGGGAGCGGATCGTAACCACCAAGCT
AGCCTGCAGCAGTGAATATCCCACCGCAGCCGGCAGAGTGTCTGCCGGTACTGAGAAATATCAGCTGG
CGTGGCAGGCCAACCGCCTACGCGCTATATATCCGCACTATGGAGCCACGTAAGGAAATTACCGAGGCCCTGCTGCATGCC
TATCAGCTGCTACCGATGGAGATCCGATGCCCTACGCGCTATCCGCTAACCTTAGCGGCTATCGGGCTGGAGAAG
CGATGTGCTGGCAGTTAGTCGAGCGCGTACTCGGGAAAGGCTGGATCCGTTACCTCGGATATGTTCCGG
ATGAAGACCTGCCGTATCTGTACGCAGCGGCCAGAGCTTGTATTCCCTCTACGAGGGATTGGT
TTACCTATTCTGAAGCGATGTCTTGGGTGCGCGTAGTATGCTCAATGTCACCTTGCCTGAGGT
TGTTGGCGATGCCGGCCTCGTTGCCGATCTTAATGATATAGACCGGATTAGCGCGCAAATTGAGG
TGCAAGATGATAGCTGGGGAAATGCCACCGCGCGGTCTGCTCAGGCGAAACAGTTTGTGGAG
AACTGTGCACACAGACCATTAACGCCATAAATTACTCTAAGGGTGTCAAGTTGAGAGTTCTACACGTCTA
TAAGACTTACTATCCCACCTACGGCGGTATTGAGCAGGTCAATTATCAGCTAAGTCAGGGCTGCC
GCCGGGGAAATCGCAGCCGATGTTTCACTTTAGCCGGACAAAGATAAGGCTCTGCTACGAAGAT
CATCGGGTCAATTATAAAACAGCTTTGAAATTGCCCTCACGCCGTTTCGCTGAAAGCGTAAAGCG
TTTAAGCTGATTAAGAGTACTACGATATCATCAACTACCAATTCCGTTCCCTTATGGATATGCTGC
ATCTTCCGGCGCCCTGACGCCAGGACTGTGGTACCTACTCTGATATAAGTGAACAAAAACGGTTA
ATGAAGCTGTACCGCCGCTGCAGGAGCGATTCTCAGCGCGTAGATTGCACTCGTTGCCCGCC
TTACGTGGCTTCCAGCCAGCCGAAATTACTCTGATAAAACGGTGGTACCCGTTGGTCTGGAGC
AGCAGGACGTGCAGCACGATCCGCAAGGGTCCGCACTGGCGGGAAACTGTCGGCATAAGTTCTTCTC
TTCGTCGGCACTTCCGCTACTACAAAGGGCTGCATATTCTGATGGATGCCGTGAGCGTAGCCGACTGCC
AGTGGTGGTTGAGGGGGGGCCGCTGGAATCGGAAGTGCAGCGTAGCGCAGCGGGCTGAGCA
ATGTGATGTTACCGCATGCTCAACGACGAAGATAAGTACATTCTCCAGCTCTGCCGGGGCGTGGTA
TTCCCTCGCATCTGCGCTCTGAGGGCTTGGCATTACGTTATTGAAAGCGCACGCTTGCAGGCC
GATCTCTGCGAGATCGGTACAGGTACCTCTTCAATTACCGAACAGTGAAGTGGTGCCTGATTCCGC
CGAATGATAGCAGCGCTGGTGGAGGCATGAATGAGCTGGAATAACGAGGAAACCTCAACCGCTAT
GGCGAAAACCGCTGCTGTTGAAAGAGATTTACTGCCGACCATATGATTGACGCCATGTCAATCT
CTACACTACATTGCTGGAAAGCAAATCTGAGCGGCCGAGCTCGCACTCGAGGATCCGTGAGGCTG
GAGCTGCTCGAAGTTCTATACCTTAGAGAAATAGGAACCTCGGAATAGGAACATAAGGAGGATATT
ATGGATAAAAGCCGTAAGCATATAAGCATGGATAAGCTATTATAACTTAAAGTACTTGTATACTTATT
TGCACATTCCAGGCCGAGCATTAGCGCGGTGATCACACCTGACAGGAGTATGAAAGTCAAGCAA
CAGATCGCGTAGTCGGTATGGCAGTGGACGCAACCTGCGCTCAACATCGAAAGCCGTGGTATA
CGTCTCTATTTCACCGCTCCCGTGGAGAACAGCGGAAGAAGTGGTACCCGAAACACTGG
TTCCTTACTACCGTAAAGAGGTTGTCGAATCTCTGAAACGCCCTCGCATCCTGTTAATGGTAAA
GCAGGGCAGGCCGAGGATGCTGCTATTGATTCCCTCAAACCATATCTGATAAAAGGAGACATCATCATTGA
TGGTGGTAACACCTTCTCCAGGACACTATTGCTCGTAATCGTGAGCTTCTGAGCGAGGGCTTAACTTCA
TCGGTACCGGGTTCTGGCGGTGAAGAGGGGGCGCTGAAAGGTCCTTCTATTATGCCCTGGTGGCCAGAAA
GAAGCCTATGAATTGGTAGCAGCGACCGATCCTGACCAAAATGCCGCCGTAGCTGAAGACGGTGAACCATGCGT
TACCTATATTGGTGGCGATGGCGCAGGTCACTATGTGAAGATGGTACAAACGGTATTGAATA
TGCAGCTGATTGCTGAAGCCTATTCTGCTTAAAGGTGGCCTGAAACCTCACCAACGAAGAACTGGCGAG
ACCTTACCGAGTGGAAATAACGGTGAACCTGAGCAGTTACTGACATCACCAAAAGATACTTCACCAA

AAAAGATGAAGACGGTAACTACCTGGTTGATGTGATCCTGGATGAAGCGGCTAACAAAGGTACCGGTTAAATGGACCAGCCAGAGCGCAGCTGGATCTCGGCAGACCGCTGCTGATTACCGAGTCTGTTGCACGTTATATCTCTCTGAAAGATCAGCGTGTGCCCATCTAAAGTTCTCTGGTCCGCAAGCACAGCCAGCAGGCGACAAGGGCTGAGTTCATCGAAAAAGTTCTCGTCGTGCGCTGTATCTGGCAAAATCGTTCTACGCCAGGGCTCTCAGCTGCCTGCAGTCAAGAGTACAACACTGGATCTGAACACTACGGCAATCGCAAGGATTTCCTGCTGGCTGCATCATCGTGCAGTCCCTGCAGAAAATCACCAGTGCCTATGCCGAAATCCACAATCGCTAACCTGTTGCTGGCTCCGTACTTCAAGCAAATTGCCGATGACTACCAGCAGGCCAGCTGCGTGTGATTCGTTGCTTATGCAGTACAGAACGGTATTCCGGTCCACCTCTCCGACGCCAGCGGTTGCCTATTACGACAGCTACCGTGCCTGCTTCTGCCATCGAACCTGATCCAGGCACAGCGTGAECTATTGGTGCCTACTTAAAGCGTATTGATAAAGAAGGGTGTCCATACCGAATGGCTGGATTAA

SEQ ID NO: 15 (example O15 *rfb* locus nucleotide sequence – O15-EPA production strain
stLMTB11738)

ATGACGAATTAAAAGCAGTTATTCTGTAGCGGGCTCGGGATGCATATGTTGCCACTAACGGCGATACCCAAAGAGATGCTACCAATCGTGACAAGCCAATGATTCACTGAGTACATTGTTGACGAGATTGTGGCTGCAGGGATCAAAGAAATCCTCTGGTAACTCACCGTCCAAGAACCGGGTCGAAAACCAACTTCGACACACTCTTATGAGTTAGAATCACTCCTTGAGCAGCGCTGAAGCGTCAACTGCTGGCGGAAGTACAGTCCATCTGTCCGCCGGCGTGAACATTATGAACGTCGTCAGGGCGAACCTTAAAGTTAGGTTAGGCCACTCCATTGTTGTGCGCAGCTGCCATTGGTGAACAACCCATTGTCGTTGACTGCCAGACGTTGTGATCGACGATGCCAGCGCCGACCCGCTACGTTACAACCTTGCTGCCATGATTGACCTTCAACGAAACGGCCGCAAGCCAGGTGCTGGCAAAACGTTATGCCGGGTGACCTCTCTGAATACTCCGTATCCAGACTAAAGAGCCGCTGGACCGTGAGGGTAAAGTCAACCGCATTGTTGAATTATCGAAAAAACCGGATCAGCCGAGACGCTGGACTCAGACATCATGGCGTAGGTCGCTATGTGCTTCTGCCGATATTGGCCGGAACGTTGACTCAGCCTGGTGATGGGACGTATTCA GCTGACTGATGCTATTGCCGAGCTGGCGAAAAAACAAATCCGTTGATGCAATGCTGATGACCGGCGACAGTTACGACTGCCGAAAAAAATGGCTATATGCAGGCCTTGTGAAGTATGGCTACGCAACCTGAAAGAAGGGGCGAAGTTCCGTAAGGTATTGAGAAGCTGTTAAGCGAATAATGAAACGTTGCAATCTGACGGATGTAACGGTTGATAAGAAAATTATAACGGCAGTGAACATTGCGAGCAAAGTAATTGTTGCGAATCTCTGCCGTTTTAATAAACCATCAGAATAACAACGAGTTAGCAGTAGGGTTTATTCAAAGTTCCAGGATTTCCCTTGTTCCAGGATTGGGATTGGTAAGACAATTAGCGTTGAATTTCGGGTTAGCGCGAGTGGGTAACGCTCGTCACATCATAGGCATGCATGCAGTGCCTGGTAGCTGTAAGCCAGGGCGGTAGCGTGCATTAAACCTCTATTAACTCAAACGAGAGCCGCTTATTCAACAGCATGCTCTGAAGTAATATGGAATAAAATTAGCTAGCATGAGC AAAACTAAACTAAATGTTCTTACCTTGCAATAAGTCAGGGTCCAATTACCTACTGCCATTATAATTTCCTTATCTGGTTAGAGTCATTGGTGATCGAATTGGTGTAGCTGAGTTTCACTGATAACTATAACAGTGTTTAATGGTTGTTGAATATGGTTGGATATAGTGGGACAAGAGAAATAGCACTAAATAACGATAAAAAATACCATTCTGAATTGGTTGGTGTGCAAGGTTGCTGCTCGTTATCTAATTATAAAACCGCAACGCCGATGTACATCAGTGCCTTATTGAGCATGCCATTGGTATTCTGATTCTGTGGCGTGCCTACTTACTTATTACAAGGAGATTTTTATGAGGGCCACCGATAAAGAAAATTCAAGTAATTAAAAATGGATTCAATTGGTTGTTGGTTCTGTCGTTAATTGCAAGGTTGCTGACTAGTGCATACACAATGTTGACCCCTTGTATTGGGTTGGCGTATCTGAAAGTTTGATGAGGATCTTAACTCAGCTAACATGATCAAACAAAGGTTGGCTGGACTTGCATCACCATTAGTCCAAAGCTTTTATCCAAGAATTAAACATTGCAAAGAGAGAATCCATATATTGCAAACACTTAAATCTAGAATGATTCTTAAACTCTGCTTGGTTTACATGGCTTCACTGGCAATACCATTGGCTTCTGATAACTTGTGCTTCACTGCTTCTGTTAATGACATTGCTTCTATATTCATAGGTTTAAACAGTTGTCGGGTTACTTGTATTAGTACCTAATGGGATGCAAAACAGTATTCAAACTCTTCTAGGAACATTACTTGTGTTAACGATAGTTATCCAGCATGTAATATTATGGAGCAACGGGTTGCGATTGTGAGTCTTGTGAGGTTATTGAGCTGAAATTTCGTTGGCATGGGATGCTTAAACAATTGCTTAAAGTAAATAAAACCGTATGTAGGCTCATAAATTATGAATATCTCGGTAAATAATATCTGTTGGAAACGCCAGTTCATTAGAATTGATTCTCTGAGCTGATTCTCAGGCTAAAGACAATAGTCTACACCTAGAAGTAATTGTTCCGATAGTCAGTGGTAAAGAAAATTGATGATGTTAGTGTGCTGATAATTCAAAAGAAAATTAA

TATTATCCATCAACATACTAAAAATATACTCTCGCTAACGCATTCGGAGCATCCCCTAGCCCCATGGGG
ATTATTTAATATTCCTGATGATGTTAGTGTATACCCGCAAGTGGATATATCATCGTTGCTGAACATTAA
AAAAAAATGAATAGTAAAAGCCTTATGTGGGAAGTAGATTGAAATGAACCTATTGAGACCACCAA
TTACTATCGCTACAGGAACCTTTACACCCCTAAGTTAGTGTAGTCCTGATATCTCTATGAATGCCCTGGA
CTTTGTCGCAATGAATTGTTCTGATAGAAAAGGCATTTCATCAGGTATAGTTCAATATAATGAAAAT
TTTATTGGTTATGGTTGTGAAGATCATGAGTTGGGTGCAACTTGAAAAAAATGACTTCAAAATTATTT
TGCTGATTTAAAATATTACATCAGAATACAGTGGCGATAGAAGGATATACAACAAATTGCTGCA
CAGCACGTGATGGTATGAATGTATTAAGCAAAGTAAGGCCTGAAATGTTCTACTAATAAAATTATTC
CTAGTTGAGAAAATATTAGTAAACACAAAAGCTTAGTAAAATATGCCAATCAATATTTCAATAAATT
TATTTTAAAAAAATAATACAATTAAAACAGATGCAAATAAAACACTTATTTCCAATTCTT
ACAGATATGTGTTGATTCGGCATATATACATGGTATTGGAGAGCGTGGCACCTCAAAACAGATGATTG
CTTAAGAACTGGTATATATAGATGATGCTATCTCATTATTAAGACATTGTTATGGAAGGTAACAA
TGAAGTATAATGCATTGATGGCTTTTATTATTTTGTGTTTTAGATTGTCGCTGATAATACCT
TTCTTATATTGGCATTATTCTGCATTGGTATTATGTATTAGTGCCTAATTGATTACTAT
GGGCAATGGATTGGTATCTAGATCGAAAATTGTTGCTTATCTATATTCTATAATTATTTTAT
TTTGTGTTGGTTTCGATTGTTCAAAAAGCATTCTTCAAAGTTATTTACCGTTAGATTATTATG
TTGTTTATTTCATTGTCCTGCGTATTAGTAAATAGATTCTAAAGGGTGACTGAAATTAA
GGAGCGAATATTAGTGTATTCTCTGGGTCAAATAGTTATTTGGTATGTATAAAGTCCAGAGT
TAAAAAGATTGTTATATACTTTCTGGTATGTCGACTCTGTTAATCTTGGAAACAAATGCTAAAGTA
AGAGGATTGGGTGTCGGGTGAAATAAAATTCTAGACACCATTGATGATCTATATGTCATTTTAT
GATGAAAAGCGTTATGCTTAATTACTTTAATTGTCGACTCAAATCGTAAATTCTAACATGGCTGTGA
TTGCAGCCATTATTGGTATCGGTTGCTTAGACCTAATATTAAATATAAAATTGCAACAGTATTGATTG
GGAGTTTAGTTATAGCTTAGGAGCGGTGTTCTTCCTCGATTATGATGAGGTTGCTTCTGGAGATGG
CACAAGAACTCTGGATATCTTATTACAGCAACATGTGTTGTTGAGTAATTAGATTTTTAATATTA
TATTGGATTACAGAAAACATATCTCATCAATCCCCGATATTAAACAAAGTCGGATATGGCTGGGTT
ATACTGTTAATTACGGGGTTAACATTATTACACTCTTTATTAAATCTTACTATTCTATTGC
GACATTGGAATGACATATCAAGCAATTATGGATGTTAATTGGATAATTCAATACCAAAAGGTTAG
TTTAGGATCTAACGGCTATTCTTCTATCTTATTATATGTTGAATAGAGTAACACTTAGTGA
CAGAGTTCAATTACTAATAAGTTAGGTCAAGTAAGTAAATAGCTTCAGAGTATATTGCAATGATTG
GGTTGGTTATTATGTTCATCTAAACACTGTTAATTACTGGTGGTACTGGCTTCTGGGAATGCTGT
ATTAAATAGATTCTGATACAGATATTGCAAGAAATCCGTATATTAGTCGTGATGAAAAAAACAAGATG
ATATGCGGAAAAAAATACAATAATCAAAATTAAAGTTCTATATTGGTATGTCAGAGATTACCGTAGTATT
TTGAATGCGACTCGCGGTGATTATATCATGCAGCGGCACTAACGCAAGTCCATCATGTGAATT
TCATCCTATGGAAGCGTAAAACATAATATCCTGGTACGGAAAATGTTCTGAAAGCAGCTATAGCGAATG
AAAGTGAAGAGGGTTGATGCTTAAGTACTGATAAAGCTGTATACCCGATTAACGCAATGGGTATTCAAA
GCTATGATGGAAAAGGTCAATGGTCGCGAAATCCGTAATGTTGATCGCAATAACAGTAATATGTTGAC
CCGTTATGGGAATGTTATGGCATCTCGCGGTTCAAGTTATCCATTATTGTTGATCTTATTAGAGCGGGCA
AGCCACTCACAAATACTGATCTTAATGACCCGCTTATGATGACTCTGAGGATGCGGTAGATTAGTT
CTTATGCGTTGAACATGGTAATAATGGTATCTTGTGCAAAAGCACCTGAGCAACTATTGACAC
ATTAGCTATTGCTTAAAGGAATTACTAAATGTTCTGACCATCCGTAATGTCATTGGAACCGTCAATG
GCGAGAAAATTATGAAGCTACTTAGTCGTGAGGAAATGATGCTGCTATAGATATGGCGATTATTAC
CGTGTCCGCCAGATCTCGTGACCTTAATTGGCAAAATGTTGAGCAAGGTGATGCCGAATATCTGA
AATAGAAGATTATAACTCTATAACTCAACGGTTAGATGTTGAAGGCATGAAAGAGCTTCTGCTAAAAT
TAGCCTTATTGAGCAATTGCGTGGTAAAAATATACTGGATTGATGATGAAATATTAGTTAC
TGGTCAAATGTTTATTGGCGTAAATTGTTGAGGCTTGAGGAACCTGGTTATAAAGATCTTATTA
GAATTGATCGAGAATCAACGAAGCAAGATCTGAACAAGGCTACAGGATGCCGATTTATTACTTA
GCTGGTATCAATAGACCTAAGACTGATGAGTTATTCTGGAAACAGTGAATTAAACAGCATATAGT
TGAGTATCTCCTTCTATTGGTAAGAATACACCAATTATGCTAACGCTTCTGATACAAGCTGAACCTAATA
ATGCTTATGGGGTTAGCAAAGCTGTAGCTGAAAGCTATGTCGAAAATATGCTGCTGCTAGTGGTTCTCG
TATTATATTCTCAGATATCCAAACGTTTTGGTAAATGGTGAAGCCAAACTATAATTCTTTATAGCAAC
TTTTGCTACAATATTCAATGATATTGAGATTACTATCAATGATGTCAGCAGCGCCAGTCATCTGGTCT
ATATTGATGATGTTGACTGATGCTATAGCTTCTCTGGGACGGTTGAAAGTGGATATAAAGTTGTT
GCACCAATTATTCAACACAGTGGTGAAGTGCAGAAATTATAGCTTCAAAATAGCCGTTCCAC
CCTGATCACAGAGGCTGCGGGGGGATTACCCGTGATTGATCTACATGGCTGAGTTATTACAG
CAGAGAAGTTGCGTACAAGGTACCTTTATGGGGATGCCCGGGAGTCTTGTGAGATGTTGAAAACG

CCTTCAGCGGGGCAGTTTCATTTTACTGCTCACCTGGTATTACGCGTGGCGGACATTACCATCACAG
 TAAAAATGAGAAGTTTGGTCATCGAGGTAGGCATGTTAAATTGAACATGTGATTACCGGTGAGC
 GATATGAACTGAAAGTTCATCGGGTGAGTTAACGATTGGTAAACAGTTCTGGTGGACACATGACATT
 ACAAAATATTGGAACTGATGAATTAAAGTCATGCTCTGGCAAATGAAATTTCACCGTGATGAGCCGA
 TACTATTGCGAGACCTCTATAATGAAAAAATTAAAGTTATGCTGTTGGAACCCGCTGAGATTAT
 CCGTTGTCGAGGGTTCTGCTAAGTTGATGAATACTGCGAGCATATTATTGCCACTGGTCAAAATT
 ATGATTACGAATTAAATGAAAGTGTCTCAATGACITGGTGTGCGAAAACCTGATTATTAAATGCA
 GCGGTAAAATGCGCGAAACCATTGGTCAGGTTATTAAAGGTAGATGAAGTATTAGAAATCGAAAA
 ACCTGAAGCAATACTGGTATTGGCGATACGAATTCTACGTTAGTATTCTGCCATTCCGCCAACGCGTAAAG
 TGCCTATATTCCATATGAAAGCAGGTAAACGTTGTTGATCAACCGTGCTGAAGAAACCAACAGACGT
 ATTGGTGACCATACGGCTGATATCAATATGACCTACAGTGATATTGCTCGTAATATCTTGGCTGAAGG
 TATCCCAGCTGATCGGATCATAAAACCTGGTAGCCCTATGTTGAGGTTCTTCATATTATGCCAAA
 TTGATGGTTAGATGTGCTATCGCGTTGAATCTACAGTCTGGTGAGTTTTGTTAGTAAGTGCATCGT
 GAAGAGAATGTTGATTCTCCAAAACAGCTCGTAAAGCTTGCACATTCTAAATACTGTTGCTGAAAATA
 TAATCTCCAGTTATTGCTCCACACACCCAAGGACACGTAACCGAATCCGTGAGCAAGGAATTGAATTTC
 ATTCAAATATAATCTACTGAAACCATTGGTTCCATGATTATAACCACCTGCAGAAGAAACTCAGAGCT
 GTGCTTCAGATAGCGGTACTATCACTGAAGAGTCATCCATCATGAATTCCCAGCGTAAACATCCGGG
 AGCGCATGAGCGTCCGGAGGCTTGAGGAAGCATCCGTATGATGGTGGGGTTAGAGTGTGAACCGCTAT
 TACAAGCGCTGGATATTCTGGCAACACAACCGCGAGGTGAAGTCCGTCTTACGTCAAGGTTAGTGAATTAC
 AGCATGCCAAATGTGCGATAAAGTTGTCAGAATTGTTCACTTACACAGATTATGTTAAGAGAGTCGT
 CTGGAAAGAATATTGATGAAACTGCTTAATCATAGATGATTACCTGCCAACAGTACTCGTGTGGTGC
 AAAAATGTTCATGAACTGCTCAAGAATTATCCAGCGTGGCACGATGTTACGTTAATTACTCCTGGTA
 CGGGCATGCAAGAAGAGATTCTTTGATACCTTTCAGGGGGTAAAACATGGCTTTAAAAGCGGGCG
 CTCAAGGATGTAAGTAAAATTCAAGCGAGCGTCAATGAAACGCTTGTCTATCGGGCTGGAAAGCCAT
 CAAAAAATGGTAAAAAGAGACCTTGAGGGGGTGATTATTACCTCCATATTCTGGGGCCTT
 TAGTAAAAAAATTAAAGCTCGTGGCAATGTCCTGCTTATCTTAAAGAGATATGTTCCACATGG
 GTAATTGATGCAAGGAAATGCTTAATGCTGGTCCCCAATAGAACGCTACTTCGTCTTTGAAAGGAAATATC
 TTATCGTCAGGCAAATCGTATTGGACTTATGTCATGATAAGAATCTTGATGTTTCGAAAGATAATAAG
 GCTATCCGTGCGAAGTTGCGTAATTGGCATCCCTAACACCAACGATCATACCAAGGATTATACCA
 CTACGTAAGCGACTTGGCTAGAGGATAAAACCATTTCTATGGTGGAAACATAGGTGATGCACAGGA
 CATGACAAACTTGATGCGACTTGTGAGAACATGGCAGCATATCCTCAAGCTCATTCTTATTGGCC
 AGGGGGATGAAAGTGAATTAAATTCAATTAGCATCTGAGTGGCATTGACGAATTTCACCTATTGCC
 TCGGTTACCAAGATGAATTAAAGTCATTTGTCGGAATGGATATCGGCTTGTCTTCTCGCTAG
 ACACCTTCCCATAATTTCCTGGTAAGTTATTAGGCTATATGGTCAGTCGCTACCTATTAGTAGCG
 TAAATGCCGAAATGATTGCTCGACATTGTCAATAAAATAATGCCGTTAATCCATGTCATGGTGG
 GACGATAATTATGTCATCTGCGCTATTAAATGTCATGATATTGATGTCGCGCCGGCAACTGGTTCGG
 GGCAGATATTGTCGAAAGACAATTCTCGGTTGAGTCGCGGCACAGACGATAGAAATGAGGTTGGAGG
 CATGCAATGCGATTAATTGATAATGACCAACTCGACGAATTATATGATCAAGCCGGCAATCGGAAACGTT
 ACGTCCCACCTTATGATGCACTGGCTCGCATCAAGAAAAGGTACAGCGTTACTATTGCAATTAGTAAGG
 GCAGCTATGTTGAAACCGCATTATCACGAACCTCTCATCAGTGGGAAATGTTCAATTGTTATGGAGGGCAA
 CTTCAAGGTTGTTGATGGTAGAAATGGTAGGTTATAAGCAATTATAGCAGGAGATAATACTGGAAT
 GAGCATTGTTGAGGTTCTCCGGGATATAACACAGTGTGCAATGCCTATCTCGCGTGTCTTATGGTGG
 AAGTTAAGGAGGGGCCATTGACCCCTTTGCAAAATGTCGTTGAGCGCCGCGAGCTCGTCACT
 CGAGGATCCGTGAGGCTGGAGCTGCTGAGTTCTATACCTTAGAGAAATAGGAACCTCGGAATAGG
 AACTAAGGAGGATATTGATGAAAGCCGTAAGCATATAAGCATGGATAAGCTATTATACTTAAATA
 AGTACTTGTATACTTATTGCAACATTCCAGGCCGAGCATTAGCGCGGTGATCACACCTGACAGGA
 GTATGTAATGTCACAGACGATCGGCGTAGCGGTATGGCAGTGATGGGACGCAACCTGCGCTCAACA
 TCGAAAGCCGTGGTTACCGCTCTATTTCACCGTCTCCGTGAGAAGACGGAAGAAGTGTGCGAA
 AATCCAGGCAAGAAACTGGTTCTTACTATACGGTAAAGAGGTTGTCGAATCTGGAAACGCCCTCG
 CATCCTGTTAATGGTGAAGCAGGGCAGGGACGGATGCTGCTATTGATTCCCTCAAACCATATCTCGATA
 AAGGAGACATCATCATTGATGGTGGTAACACCTTCTCCAGGACACTATTGTCGTAATCGTGAGCTTCA
 GCAGAGGGCTTAACTCATCGGTACCGGTGTTCTGGCGGTGAAGAGGGGGCGTAAAGGTCTTCTAT
 TATGCCCTGGTGGCCAGAAAGAAGCCTATGAATTGGTAGCACCACGATCTGACCAAAATGCCCGT
 AAGACGGTGAACCATGCGTACCTATTGGTGGCGATGGCAGGTCAGTGTGAAGATGGTCAAC
 GGTATTGAAATACGGCGATATGCAAGCTGATTGCTGAAGCCTATTCTGCTTAAAGGTGGCCTGAACCTCAC

CAACGAAGAACCTGGCGCAGACCTTACCGAGTGGAAATAACGGTGAAC TGAGCAGTTACCTGATCGACATCA
 CCAAAGATATCTCACCAAAAAAGATGAAGACGGTAAC TACCTGGTTGATGTGATCCTGGATGAAGCGGCT
 AACAAAGGTACCGTAAATGGACCAGCCAGAGCGCCTGGATCTGGCAACCGCTGCGCTGATTACCGA
 GTCTGTGTTGCACGTTATCTCTCTGAAAGATCAGCGTGTGCCCATCTAAAGTCTCTGGTC
 CGCAAGCACAGCCAGCAGGGCTGAGTCATCGAAAAGTCGCTGCGCTGATCTGGCAAA
 ATCGTTCTACGCCAGGGCTCTCTCAGCTGCGTGTGCGCTGAAAGAGTACAACACTGGGATCTGAACTA
 CGCGAAATCGCGAAGATTCCGCTGCGTGCATCCGTGCGCAGTCCAGAAAATCACCAGAT
 CTTATGCCGAAATCCACAGATCGCTAACCTGTTGCTGGCTCCGTACTTCAAGCAAATTGCCGATGACTAC
 CAGCAGCGCTGCGTGTGCGTGTGCTGCTGCGTGCACAGAACGGTATTCCGGTCCGACCTTCTCCGCAGC
 GGTTGCCTATTACGACAGCTACCGTGTGCTGCTGCGTGCACAGCGTGAATT
 TTGGTGCGCATACTTATAAGCGTATTGATAAAGAAGGTGTGTTCCATACCGAATGGCTGGATTAA

SEQ ID NO: 16 (example O16 *rfb* locus nucleotide sequence – O16-EPA production strain
stLMTB11739)

ATGACGAATTAAAAGCAGTTATCCTGTAGCGGGTCTCGGGATGCATATGTTGCCTGCCACTAAGGCAT
 ACCCAAAGAGATGCTACCAATCGTCACAAGCCAATGATTCACTGAGATTGACGAGATTGTGGCTGCAG
 GGATCAAAGAAATCCTCTGGTAACTCACCGTCCAAGAACGCGGTCGAAAACCACTTCGACACCTCTTAT
 GAGTTAGAATCACTCCTTGAGCAGCGCGTGAAGCGTCAACTGCTGGCGGAAGTACAGTCCATCTGTCCGCC
 GGGCGTGAACATTATGAACGTGCGTCAGGGCGAACCTTAGGTTAGGCCACTCCATTGTGTGCGCAG
 CTGCCATTGGTACAACCCATTGCGTGTGACTGCCAGACGTTGATCGACGATGCCAGCGCAGCCGACCCG
 CTACGTTACAACCTTGCGCATGATTGACGTTCAACGAAACGGGCGCAGCCAGGTGCTGGCAAAACG
 TATGCCGGGTGACCTCTGAAATACTCCGTACCCAGACTAAAGAGCCGCTGGACCGTGAGGGTAAAGTC
 GCCGCATTGTTGAATTATCGAAAACCGGATCAGCGCAGACGCTGGACTCAGACATCATGGCGTAGGT
 CGCTATGTGTTCTGCCATATTGGCGGAACCTGGAACTGGAACGTAACGCTGGCGATGGGACGTATTCA
 GCTGACTGATGCTATTGCGAGCTGGCGAAAAAACAAATCCGTTGATGCAATGCTGATGACCGGGCAGAGTT
 ACGACTGCGGAAAAAAATGGCTATATGCAGCGTGTGAAAGTATGGCTACGCAACCTGAAAGAAGGG
 GCGAAGTCCGTAAAGGTATTGAGAACGCTGTTAAGCGAAATAATGAAAATCTGACCGGATGTAACGGTTGAT
 AAGAAAATTATAACGGCAGTGAAAATTGCGAGCAAAAGTAATTGTTGCGAATCTCCTGCCGTTTTA
 TATAAACCATCAGAATAACAACGAGTTAGCAGTAGGGTTTATTCAAAGTTTCCAGGATTTCCTTGTT
 CCAGAGCGGATTGTAAGACAATTAGCGTTGAATTTCGGGTTAGCGCAGTGTTAACGCTCGTCAC
 ATCATAGGCATGCAGTGCAGTGTCTGGTAGCTGAAAGCCAGGGCGGTAGCGTGCATTAAACCTCTATT
 AATCAAACGTAGAGCCGCTTATTACAGCATGCTCTGAAGTAATATGGAATAAAATTAAAGTAAAATCTT
 GTTACTGGGGCGCAGGATTATTGGTTCAGCTGTAGTCGTACATTATAAAATAACGCAGGATAGTGT
 TGTAAATGTCGATAAAATTAAACGTACGCCGAAACCGGGAAACTCTGGCTGATGTTCTGATTCTGAAAGCCT
 ATGTTTGAACATGCCGATATTGCGATGCACCTGCAATTGGCACGGATTGGCTCAGCATGCCGGAT
 GCAGTGATGCACCTGGCTGCGTAAAGCCATGTTGACCGTCAATTACAGGCCCTGCGGCAATTATTGAAAC
 CAATATTGTTGTTACTTATGTCCTTTGGAAGCCGCTGCAATTACTGGTCTGCTCTGATAGCGACAAGA
 AAAATAGCTCCGTTTCATCATATTCTACTGACGAAGTCTATGGTATTGCGCTCATCCAGATGAAGTA
 AATAATACAGAAGAATTACCCATTATTACTGAGAACGACAGCTACGCGCCAAGCAGCCATTACCGC
 CAAAGCATCCAGCGATCATTAGTCGCCGCGTGGAAACGTACATATTGTTACCGACAATTGACTAATT
 GCTCGAACAACTATGGTCCTTATCATTCCCGAAAAGCTTATTCCACTGGTTATTCTAATGCACTGGAA
 GGTAAGGCATTACCTATTATGGCAAAAGGAGATCAGATCCGCGACTGGTTGATGTTGAAGATCATGCGCG
 TGCCTTATATACCGCGTAACCGAAGGTAAAGCGGGTAAACTTATAACATTGGTGGGCACAACGAAAAGA
 AAAACATCGATGTTGCTCACTATTGTTGATGCTGGATGAGATTGACCGAAAGAGAAAATCTTATCGT
 GAGCAAATCACTTATGTTGCTGATCGTCCGGGACACGATCGCCGCTATGCTATTGATGCTGAGAAGATTGG
 TCGCGCATTGGGATGGAACACAGGAAACGTTGAGAGCGGGATTGTAACCGGTGGAATGGTACCTGT
 CCAATACAAAATGGGTTGATAATGTAAGAAAAGTGGTGCCTATCAATGTTGAGATTGAAACAGAACTATGAGGGC
 CGCCAGTAATGAATATCCTCCTTTGGCAAAACAGGGCAGGTAGGGAAACTACAGCGTGTCTGGCA
 CCTTGGGTAATTGATTGCTTGTGATGTTCACTACTGATTATTGCGGTGATTGTTAGTAATCTGAAGG
 TGTAGCTGAAACCGTAAGAAGCATTGGCGGATATTATGTCATGCAATGCAAGCCGCTCACACCGCAGTAGACA
 AAGCAGAATCAGAACCGGAGTTGCGACAATTAAACGCAACAAGTGTGCAAGCGATTGCGAAAGCAGCA

AATGAAGTTGGAGCCTGGTTATCCATTACTCGACTGATTACGTCTCCCTGGAAATGGCGATATGCCATG
 GCTGGAGACGGATGCAACCGCACCAATAATGTTACGGTAAACCAAGTAGCCGGAGAAAAAGCGTTAC
 AGGAATATTGCGCGAAGCATCTTATTTCCGGACCAGCTGGGTCTATGCAGGAAAAGGAATAACTTCGCC
 AAAACGATGTTACGTCTGGCAAAGAGCGTGAAGAATTAGCGGTTATTAACGATCAGTTGGTGCGCCAAC
 AGGTGCTGAACGTCTGGTGATTGTACAGCACATGCCATTGCGTGCACGTGAATAAACCGGATGTCGAG
 GCTTGTAACATTGGTAGGCCAGTGGTACCAACCTGGTACGATTATGCTGCCTGGTTTGAAGAGGCG
 CGAAAGCAGGCACTCCCTGCACTCAACAAGCTAACGCAACAGCCTACACCAG
 TCGTCGCCACATACTCTGCCCTAACAGAAAATTTCAGCAGAACTTGCCTGCTTGCCTGACT
 GGCAGGTTGGCGTGAACGAATGCTCAATGAATTATTCAGACTACAGCAATTAAATAGTTTGATCTT
 GTTGTGATGGTGGAGCAAGATGAATTAAAGGAATGATGAAATGAAATGCGTAAAGGTATTATTTAGC
 GGGTGGTCTGGTACACGCTTTATCCTGTGACTATGGCTGTCAGTAACAGCTATTACCTATTATGATA
 AACCGATGATCTATTACCGCTCTCACACTGATGTTGGCGGGTATTCGCGATATTTGATTATCAGTACA
 CCTCAGGATACTCCTCGTTCAACAATTGCTGGGTGACGGTAGCCAGTGGGCGTGAATCTCAGTACAA
 AGTGAACCTAGCCCAGATGGCCTCGCGAGGCATTATCATCGGTGAAGAGTTATTGGTGGTATGATT
 GTGCTTGGTCTGGTATAATATCTTACGGTCAGATCTGCCAGCTATGGTGTGAGTTGA
 AAAGAAAGTGGTCAACGGTATTGCTATCACGTTAATGATCCAGAACGCTATGGTGTGTTGAGTTGA
 TAAAAACGGTACGGCAATCAGTCTGGAAGAAAACCGTTAGAACCAAAGAGTAATTACGCCGTACAGGTC
 TGTACTTTATGATAACGACGTGGTCAGATGGCAGGGACGCTGTCTGCGGATGATGGGCGTGGCTACGC
 GTGGCTGGACACGGGACTCATCAGAGTCTGATAGAACAGTAATTACGCCGTACAGGTC
 AGGGATTGAAGGTTCTGCTCTGAAGAGATTGCATTGTAAGGTTTATTGATGTTGAGCAAGTAAGA
 AAATTAGCTTACCAACTAAAGAATAATTATGGCAGTATCTTATAAAATGACGAAGGATTCAAATTA
 ATGAATGTGATTAGAACTGAAATTGAAGATGTGTAATTCTGGAGCCAAGAGTATTGGTGTGATAGAGG
 TTTCTTATGAGAGCTTAATCAATCAGCATTGAACATATTCTAGGCTATCCGGTCAGCTTGTCAAG
 ACAATCACTCACGTTCATCAAAATGTAACAGGCTTCACCTCAACCGGGCAGTACGCACAAGAT
 AAACCTGTACGCTGCACTCATGGAGCAGTTTGATGTTGTTGATATTGACCCAAATCGGTATCCTT
 TGGTAAATGGGTTGGTCTGCTTACGCTGATAATAAGCAGCAGTGTGGATACCAAAAGGTTGCTC
 ATGGCTTTGGTCTGCTGATATCGCTGAATTCAATATAAAACTACAAACTATTATCATCCTGAAAGC
 GATTGTGAATATGTTGAATGAAACGATGCAATTGATTGGCCCCAACATCAGGGTTAATCCTT
 GCCAAAAGATGAAAGGCTTTACGTTAGATGAGCTTACGATTAAATTAATTGATGAAACAGAATAA
 ATTATCTTAAGAAGAACGTTATATATCTGGCTGCTCAAGGTAGCAATTATCTTACCATGCTTA
 CATTCCATATCTGTAAGAACACTGGCTCTGAAAATTGCGTATTGCGTTTGTCAAGCGACTATG
 CTATATATGATAATGTTGTTGAATATGGTTCAATCTCACAGCAACTCAGAGTATTGCCAAAGCAGCAGA
 TAGTAAAGATAAAAGTAACGCTATTGGGGGGTGTGATATTTCAAAAATAGTTCTATCGTCATTACAT
 TGATTTCCTAACGTCGATGACCTGCTTGTCTGAATATAACAAGCATGCCGAATTATATGGTGT
 GTTCTGCAATTAGTCGGAATTAACTACCCATTCTGGCTGTTCAAGGGAAAAGAAAAATGAAATGGCT
 GACTTTAAGTAGTATTGTTCCCGCTGGCTATTATCCCTCAACATTATTGGTGAACACAAAGTCAG
 ATATAGCAATTGCCGGTTTATTCACTGCAAGTGCACATTGGTTGCTGGAATTATTGCACTAGCTATCGTT
 GTTCATGAAGGGTGGATTGGTAAAGTTACGCTATCATTACATAATGTCGTCGATCTTAGCAGACGGTT
 TCATGTTTATTCCACATCTGCTATTAGTTATATTCTACGGGAAATAGTTATTATCCTGGGATTAT
 CTGGACCAACGTCGCTGGGAATTAAATGCCCAAAACTATAAGAACGCGCTCAAGGGCTATTAAAT
 CCTATCACCCAAAGCAATAACCCAAAGAATATCAAGTACGCTTCTTAATCGTGTGAAGGGTGTGATT
 AATTAAGGAAATCATTGACCTGCTGAGTTGATTGGTGGTCTTTCATTAATTCTGCTTGGGTGCAT
 CTATACTAGTAAAGTATAGGCCGGATATGATAATGCACTGAGTGTGCTAATGATTATCGCCT
 CTGCCCTTCTTATTCAAGTAATGTCATGGCATTCAAGTTATGCTGACCCATAATTATAAGAAAGA
 ATTCACTAAGATTAAATCGCTGCCGGTTGTGAGTTGTTGATTTCGCTAACACTCTTTTA
 AAGAGATTGGTGCAGCAATAACATTGCTTGCACAGAGTGTCTAGTTACGTCACTCATGCTGATGTCGA
 AGAAATAAAATTACTGGTTGCTGAGGATTATGTACGATTATCATGGTTCTGGTTGTTG
 GTGCCGTTGCGAATGAGTTAAAAAGCTAAACAAAAAGCTTATTAGTGTGAGGAAAAGAAATCATATC
 GGTGGAAATGCGTACACAGAGGACTGTGAGGGTATCCAGATTCAAAATATGGTGCACATATTTCTAC
 CAATGATAAAATATGGGATTACGTTAATGATTAGTGAATTAAATCGTTTACTAATTCTCCACTGG
 CGATTATAAGACAAATTTCACCTCCTTTAATATGAATACCTCCACCAATGTGGGAGTTAAA
 GATCCTCAAGAACGCTAAATATCATTAAATGCTCAGAAAAAAAGTATGGTGCACAGGTACCTGAAATT
 GGAGGAGCAGGCGATTTCATTAGTTGGGAGGACTTACCAAGCATTGATAAAGGGTTACGGAGAAC
 AGTGGGAGAAGTGCACAAAGAATTGCCCTGCATTATTAAAGCGAATCCCAGTGAAGATTACGTTGAT

AACAAATTATTTCCGATCGCTATCAAGGTATTCCGGTGGGAGGCTACACTAAGCTTATTGAAAAATGCT
 TGAAGGTGGACGTAAGGATTAGGCATTGATCAGTACTTCGACTATAGGTTGGAGCGTTAGAATATCGCTTTA
 GAATCATCTACACTGGACCCATTGATCAGTACTTCGACTATAGGTTGGAGCGTTAGAATATCGCTTTA
 AAATTGAGACGGAACGCCATGAATTCCAAACTTCCAAGGGATGCAGTAATAAATTCACTGATGCTAA
 TGTACCATATACCAGAATAATTGAGCATAAACATTGACTATGTTGAGACAAAGCATACTGGTTACAA
 AAGAATATCCATTAGAGTGGAAAGTTGGCGACGAACCTACTATCCAGTTAATGATAATAAAACATGGAG
 CTTTTAAGAAATATAGAGGTTAGCTAGCAGAGAACAGGTTATATTGGCGGGCGTTGGCCGAGTA
 TAAATATTATGATATGCATCAAGTGTATCTGCCGCTTTATCAAGTAAAAAATATAATGAGTACGGATT
 AATGATCTATCTTGTAAATTAGTGTCTTCTATTACAGCATTATCTGTTATATCTTAAGAAGGATATAT
 TTTATCCAGCGTATGCCTTAATATCATCTTCCGACTGGCTTATTGGGATATGAAATAACGTCAGATATA
 TATGCTTTCAAGTAAATGACGCTACGTTGATTTCTACTTGCAATGTTGACATTACCTGTCTAG
 TTTATTGACGGAAGGTGATTAGATCTAAATATCAGAAAAGTCATAATGCTATTATAGCATACCATCGA
 AGAAAGTGCATAATGTAGGCTTAGTTAGTTATTCTTTGATGATATATATGATGAGGTTAAGTAAC
 TACCAAGTTCGGGACTAGCTTACTTAGCTATATGAATTGATAAGAGATGCTGATGTTGAAGACACATCAAG
 AAATTCTCAGCATACATGCAGCCATCTAACTACTACCTTGCTTATTATTTGGCTAAAAAATTAA
 CTAATACAAAGTAAGTAAACATTACTTACTTGTATTGATTATCATCTTGCATTCTGAGGTTAATAGAGTAAACA
 ACTGGTAAGCAAATTGCTTTATGGTTATCATTCTTATGCTTAGGCTGATGGGCTATTAACAGGTGG
 AGTCTCTATGTCGTTGCATAAAGAATTGTTGGGTTGGGTTGCCAACAAATGTTACTGCTTTCCG
 ATTATGTTATTTCCGCGGAGCTAAGCTATTGATGATGGTTATTCATGGCTGATTTCAGGTGTTTA
 TGGAGATTGTCGAAATTACATATCTGTGAAAATATTATTCATATTATACCTTCTTCAT
 TTTTATCATGAAAGCTTCATGACTAATATTAGCAGTTGGATACAAATACTCTTGTATCATAGTATTCT
 CTCAATTCTTAAGGCCAGAAAATAAAGTGAAATTGTTGAATGATTAAATTCTCTAGACGCG
 ATGCTGGATTAAAGCAAGAAAAGATGCACTGGACATTGCTCAGATTGAAACATTCTGTTGTTAAC
 ATTCTCTATGGGTTGGAGTAGTCCAGAGAATTATTAGTCTGTTAAGCTTAGTACATTCTCTGCGGTCT
 TGAAAATAAAGATGTTAATTCAATTCCGATGGCAAACCAATTGGCATATATTGTCATTCTTC
 ACCGCCCTCTAAATTAGAATAGTACCTCTGATTGATATTGATGAATTAGGAGGAGGGGGTAGT
 GATTCTGTGCGGCTTGCTACCTGTGATATGGTCATAAGTCACAATCCACAAATGACAAAGTACCTTAGTAA
 ATATATGTCAGGATAAAATCAAAGACATAAAATATTGATTACCTCGTCTCATCTGATGTTGGAGCATC
 GAGATGTTACGGATAAGCAACGAGGGTCATATGCTGGCAACCTTCTAGGCATAAATGTTCTTCATA
 TAACTGAAGGATGCGATTACTCTTGGTCAACTATGAAAATAAGATAATCCTAAATATCTGG
 AAGTTTGATGCTCAATCTCCGAAAAGATTAACCTCCCAGGCATGCAATTGGACTCATTGGATGGAG
 ATTCTGCGAACCTGTAGTGGCCTTGGCAGTATTAAAGTTAATAACCCCTCATAGACATCTT
 TATCTTCAATGAACTTCCAGTATTATATGGATAAAGCCGCCCCGTTGCGGATTTCATTGAGATAATAG
 AATAGGATATGCACTGGGATCAATCAAAGAAAATGCAAGAGATTGTTGACTCCATGACAATAGAAAATTATA
 AGCAAATTAGTGAGAATACAAAATTATTCTCAGAAAATTGCAACAGGAAGTTACTTCAGGGATGTTCT
 GAAGAGGTGATCGATGATCTAAAACCTCGCTAACGATATGGTCTCTGGTTTATTGGCTGTTAGAG
 ATGCTTATTGACTCGTGTATTACCGGAACTGTAGAATTATTGATTTCCCTGCTATATTGCAATGAT
 GGTAGCATTAAATTGGTAAAATTCAAGTGGAGTCGGTCTCAGGCTGGATGCTTGGACGTTGGCGT
 GATTTTTTCCGATAATGTGCAAGTTAACGACTATGTCATATGCCCTCAATTGAGAGCGTTACGATAG
 GTCGGGATACGTTATTGCAAGTAAAGTATTACCGATCATAATCACGGTCTTAAGCACTCTGAT
 CCAATGAGTCGCCAAATATACCTCAGACATGCGCACGTTGGAATTTCAGCTGTTGTAATTGGCCAGAG
 GGTTGGTTGGGAGAATGTGACGGTTTGCTGGAAACATTATTGTAATGGAGTCGTAGTCGGCGCCA
 ATTCTGTTAGAGGTTCTATTCCGAAAATACTGTCATTGCGGAGTACAGCAAAATCATAAAGAAA
 TACAATCATGAGACCAAATTATGGAAAAGCATACTGCTGTTCTGCGGTCAATTACCACTGGCGGT
 CCATTACATTGAAAAATTTCAGCAACTAATAAAAGAAAATGTCAGTTATCGCATTAGT
 CCATTCTGCTAAAGAGTTAAAGAAAAGTTATCCATGGGTTAAATTGAGTTCTGAGGTTAAAGGGT
 CGTGGCTAAACGTTGCACCTTGAATATGTTGAGTTGTTAAAGGAGCTGAATGCTACGCAT
 TGGATTGTCGATGATATTACGCCAATGTCGTCACTAAAAAGATATGTTGATTCATACCCCTGC
 CCCTTTATAAGGAATTTCATTCCGTAATTCTTATGGAGCCTAGCTTTCTTATTAAATGCTAT
 ACGGGCTGATATATAAAACATAAAAAAACTGCACTGTTGTTCAACAATTCTGGATGAAAGAA
 AAATTATCAAGAAATATCTATAAAACATCATTGTCAGTCGGCCAGAAATTAAATTATCTGATAAAAG
 CCAACTACTGATGATGATTCTCAATTAGAATAACCCCTCTGAGTTGACAATTTCACCCTGCTGTT

CACGAGTATTTAAAATTACGAGCTTATTATTAGTCAGCAAGGAAATTGAAAGAACATCCAATATTAAA
 TTTCTGCTTACTATCAGTGGTACAGAAAATCGTATGCAAATATATTAGTCAGTCTGCAGAAGGACTGGA
 TAATGTTCATTCCTCGGGTACTTGATAAAGAAAAATCGATCATGTTATAATATTAGTCAGATATAGTT
 GTTTCCCTCTAGGTAGAAACATGGGGATTGCCGTTCTGAGGCTAAAGAGCGAGGTAAAGTGGTATTA
 GCATCAGATTTCCCATTACTAGAGAAAATCTGGTAGTTATGAAAGAAAGCTTTTTGATTCTAATAA
 CGATGACATGTTAGTTAAACTTATTAGTACTCAAAAAGTAACCTCAAAAAGATATCTGTGATGAA
 ATTCATTTATCGTAATGAAATGTATTAGTGGGTTGATGAACACTAGTTAATTATTACTGAAGAACAT
 TGAAATGGTATATATAATAATCGTTCACGACATGAAGACTACATCAAAAAGTACCTCTATTGAAACAATCTA
 ATGCTGACGATGAGCACTACAAGATTACGTACCGACAAACAAAGACTCTATTGAAACAATATGC
 CAGCATTATGCAGGCTGGACTATATTAGTGGAGGTGTATACGGCTTGGTCATAATAATAATTGCGGT
 GGCATGTAAGGAAAATAGACCCGCAGATGATGATTACATTGTTGAATCCGATATCATCA
 TGAAGCATGATGATTGCTGACATATATTAAATATGTCGAAAGTAAGCGTTATGCTTTAGTACATTATGC
 CTGTTCCGAGATGAAGCGAAATCTTACATGATTATTCCGTAAGAAAATTCTGTGCTTCTGATT
 TGTGTCATTTATGTTAGGGATTAATAAAACAAAATTCTAAAGAAAGTATCTATTCTGATACGGTTGTTG
 ATTGGTGCAGGATCATTTATGCTGGTACGTTTCAGATTGTCGTTAGGCTTAGCCTGGCTGGTCAGACTTCATT
 TACTTATGTAAGATATTGACCTGTGCTTGGAGGCTTAGCCTGGCTGGTCAGACTTCATT
 TCCCGTTTCATGCGATACATTATGCTCATGACAATCGAAGTTTTCAAAAGCCTCAGATGGC
 ACTTAAAAGTACTTTAGATATTAGCCAGAAAACGTATTTATCAAATCGAACACTTGATCGAATT
 TCAGTTTCACCGTAAGAGCTCGTACCCGGGCCTAGGGTAGGGCTGGAGCTGCTCGAACGTT
 ACTTTCTAGAGAATAGGAACCTCGGAATAGGAACTAAGGAGGATATTCAATCCGTCGACGGCGGCC
 TGCAGGCATGCAAGCTGATCCATGGATCGTAGCTTAATTAAAGCGTAAGCATATAAGCATGG
 ATAAGCTATTATACTTAAAGTACTTTGATACTTATTGCAACATTCCAGGCCGAGCATT
 CGCGTGATCACACCTGACAGGAGTATGTAATGTCAGAACAGATCGCGTAGTCGGTATGGCAGTGAT
 GGACGCAACCTGCGCTCAACATCGAAAGCCGTGGTTACCGTCTCTATTCAACCGTCCGAG
 GACGGAAGAAGTGAATGCGAAAATCCAGGCAAGAAACTGGTTCTACTATACGGTGAAAGAGTT
 AATCTCTGGAAACGCCCTCGCATCCTGTTAATGGTAGAACAGCAGGTGCAGGCACGGATGCT
 TCCCTCAAACCATATCTGATAAAGGAGACATCATGTTGAGCTTACCTTCTCAGGACACT
 TCGTCGTAATCGTGAGCTTCAGCAGAGGGCTTAACTTCATCGGTACGGGTGTTCTGGCG
 GGGCGCTGAAAGGTCTCTATTATGCCCTGGTGGCCAGAAAAGAAGCCTATGAATTGGTAG
 ACCAAAATGCCCGCTAGCTGAAGACGGTAAACCATGCGTTACCTATATTGGGCC
 CTATGTGAAGATGGTTACAACGGTATTGAATACGGCGATATGCAGCTGATTGCTGAAGC
 TTAAAGGTGGCTGAACCTCACCAACGAAGAACTGGCGCAGACCTTACCGAGTGA
 AGCAGTTACCTGATCGACATACCAAAAGATATTCACCAAAAAGATGAAGACGGTA
 ACTACCTGGTTGA
 TGTGATCCTGGATGAAGCGGCTAACAAAGGTACGGTAATGGACCAGCCAG
 AGCGCGCTGGATCTCGCG
 AACCGCTGTCGCTGATTACCGAGTCTGTGTTGCACGTTATCTCTCTGAAAGAT
 CAGCGTGGCG
 GCATCTAAAGTCTCTCTGGTCCGCAAGCACAGCCAGCAGGCAG
 TCGTGCCTGTATCTGGGAAAATCGTTTACGCCAGGGCTCTCAGCTCGTGC
 AGAGTACACTGGGATCTGA
 AGTACAACCTGGGATCTGAACACTACGGCGAAATCGCGAAGATTT
 CCCTGCTGGCTGC
 CATCATCCGTGCGCAG
 TTCCTGCAAAAATACCGATGCTTATGCCAAAATCCACAGATCG
 TAACCTGTTGCTGGCTCGTACTT
 CAAGCAATTGCCGATGACTACCCAGCAGCGCTGCGTATGTC
 GTGCTTATGCAGTACAGAACGGTATT
 CGGTCCGACCTCTCCGCA
 CGGGTGCCTATTACGACAGCTACCGTGC
 CTGCTGTTCTGCCTGCGAACCTG
 ATCCAGGCACAGCGTGA
 CTATTGTTGGTGC
 GCAACTTAAAGCGTATTGATAAAGAAGGTGTGTTCC
 ATAC
 CGAATGGCTGGATTAA

SEQ ID NO: 17 (example O18A *rfb* locus nucleotide sequence – O18A-EPA production strain BVEC-L-00559)

ATGACGAATTAAAAGCAGTTATTCTGTAGCGGGTCTCGGGATGCATATGTTGCCTGCCACTAAGGGC
 ATACCCAAAGAGATGCTACCAATCGTGCACAAGCCAATGATTCAGTACATTGTTGACGAGATTG
 GGCTGCAGGGATCTGA
 GGATCAAAGAAATCTCTGGTAACCTACCGTCCAAAGAACCGGGTC
 AAAACCACTTCGACACCTCTTAT
 GAGTTAGAATCACTCCCTGAGCAGCGCTGAAGCGTCAACTG
 CTGGCGGAAGTACAGTCCATCTGTCCGCC
 GGGCGTGA
 CACCATTATGAACGTGCGTCA
 GAGGGCAACCTT
 TAGGTTAGGCC
 ACTCCATT
 TGTGCG
 GAC
 CTGCCATTGGTACA
 ACCATTGTC
 GCGTGTACTGCCAG
 ACCTGGTGT
 GATCGACGATGCC
 AGCGCC
 GACCCG
 CTACGTTACA
 ACCTTGCTGCC
 ATG
 GCACGTT
 CAACG
 GAAACGGGCC
 GAGCC
 CAGGTG
 CTGGCA
 AAAACG

TATGCCGGGTGACCTCTCTGAATACTCCGTATCCAGACTAAAGAGCCGCTGGACCGTGAGGGTAAAGTCAGCCGATTGTAATTATCGAAAACCGGATCAGCCGAGACGCTGGACTCAGACATCATGGCGTAGGT CGCTATGTGTTCTGCCGATATTGGCCGGAACCTGGAACTGGAACGTAACAGCCTGGTGATGCCGACAGTT GCTGACTGATGCTATTGCCGAGCTGGCAAAAAACAAATCCGTTGATGCAATGCTGATGACCAGGACAGTT ACGACTGCAGCAAAACAAATGGGCTATATGCAGCGTTGTAAGTATGCCCTACGCAACCTGAAAGAAGGG GCGAAGTCCGTAAAGGTATTGAGAAGCTGTTAACGAAATAATGAAAATCTGACCGGATGTAACGGTTGAT AAGAAAATTATAACGGCAGTGAAAATTGCGAGCAGCAAAGTAATTGTCGAATCTCCTGCCGTTGTTA TATAAACCATCAGAATAACACGAGTTAGCAGTAGGGTTTATTCAAGTTCCAGGATTTCCCTGTT CCAGAGCGGATTGTAAGACAATTAGCAGTTGAATTTCGGGTTAGCGCAGTGGGTAACGCTCGTCAC ATCATAGGCATGCAGTGCCTGGTAGCTGTAAGCCAGGGCGGTAGCGTCATTAACCTCTATT AATCAAACGTAGAGAGCCGCTTATTCACAGCATGCTGTAAGTAATATGGAATAAAATTAAAGTAAAATAC TTGTTACTGGTGGCGCAGGATTATTGTTAGCTGAGCTGTTGACGTCACATTATAAAATAACGCAAGGATAGTGT TGTTAATGTCGATAAAATTAAACGTACGCCGAAACCGGGAACTTGCTGATGTTCTGATTCTGAAACGCT ATGTTTTGAAACATGCGGATATTGCGATGCACTGCAATGGCACGGATTGCTCAGCATCAGCCGGAT GCAGTGATGCACCTGGCTGTAAGCCATGTTGACGTCATTACAGGCCCTGCGGCAATTGAAAC CAATATTGTTGTACTTATGTCCTTTGGAAGCCGCTGCAATTACTGGTCTGCTCTGATAGCGACAAGA AAAATAGCTCCGTTTCATCATATTCTACTGACGAAGTCTATGGTATTGCTCATCCAGATGAAGTA AATAATACAGAAGAATTACCCATTACTGAGACGACAGCTACGCAAGCAGGCCATTCCGCATC CAAAGCATCCAGCGATCATTAGTCGCGCTGGAAACGTCACATATGGTTACCGACAATTGTAATT GCTCGAACAACTATGGTCCTTATCATTCCCAGAAAGCTTATTCCACTGGTATTCTTAATGCACTGGAA GGTAAGGCATTACCTATTATGGCAAAAGGAGATCAGATCCGCACTGGTTGATGTTGAAGATCATGCGCG TGCCTTATATACCGCTAACCGAAGGTAACGGGTTGAAACTTATAACATTGGTGGGCACAACGAAAAGA AAAACATCGATGTTAGTGCCTACTATTGTTGATGCTGGATGAGATTGACGAAAGAGAAAATCTATCGT GAGCAAATCACTTATGTTGCTGATCGTCCGGGACACGATCGCCGCTATGCTATTGATGCTGAGAAGATTGG TCGCGCATTGGGATGGAACACAGGAAACGTTGAGAGCGGGATTGTAACCGGTTGGAATGGTACCTGT CCAATACAAAATGGGTTGATAATGTGAAAAGTGGTGCCTATCAATCGTGGATTGAAACAGAACTATGAGGG CGCCAGTAATGAATATCCTCTTTGGAAAACAGGGCAGGTAGGGTGGAAACTACAGCGTCTGGCA CCTTGGGTAATTGATTGCTTGTACTGATTATGCGGTGATTGTAATCCTGAAGG TGTAGCTGAAACCGTAAGAAGCATTGGCCGGATATTATGTCATGCAATGCAAGCCGCTCACACCGCAGTAGACA AACGAGAATCAGAACCGGAGTTGCACAATTAAACGCAACAAGTGTGAAAGCATTGCGAAAGCAGCA AATGAAGTGGAGCCTGGGTATCCATTACTGACTGATTACGTCTCCCTGGAAATGGCGATATGCCATG GCTGGAGACGGATGCAACCGCACCCTAAATGTTACGGTAAACCAAGTTAGCCGGAGAAAAGCGTTAC AGGAATATTGCGCAAGCATCTTATTCCGGACCAGCTGGCTATGCAAGAAAAGGAATAACTTCGCC AAAACGATGTTACGCTGCAAAGAGCGTGAAGAATTAGCGGTTATTAACGATCAGTTGGTGCCTAAC AGGTGCTGAACGCTGGCTGATTGTCAGCAGCATGCCATTGCGACTGAATAACCGGATGTCGAG GCTTGTACATTGGTAGCCAGTGGTACGACCAACCTGGTACGATTGCTGCGCTGGTTTGAAGAGGG CGAAAGCAGGGATTCCCTGCACTCAACAAGCTAACGCAAGCTACCAACACGCCATTCCACACCAGC TCGTCGCCACATAACTCTGCCCTAACAGAAAATTTCAGCAGAACTTGCCTGCTGCCTGACT GGCAGGTTGGCGTGAACGAATGCTCAATGAATTATTCAGACTACAGCAATTAAAGGTATTGATA ATTGCGTGTGGAGCAAGATGAATTAAAAGGAATGATGAAATGAAATGCTAAAGGTATTGATA AACCGATGATCTTACCCGCTCTCACACTGATGTTGGCGGGTATCGCGATATTGATTACGATACA CCTCAGGATACTCCTGTTCAACAATTGCTGGGTGACGGTAGCCAGTGGGCTGAATCTCAGTACAA AGTGCACCTAGCCCAGATGGCCTCGCAGGCATTATCATCGGTGAAGAGTTATTGGTGTGATGATT GTGCTTGGTTCTGGTACACGTTTATCCGTGACTATGGCTGTCAGTAAACAGCTATTACCTATTGATA AAAGAAAGTGGTCAACGGTATTGCTATCACGTTAATGATCCAGAACGCTATGGTGTGTTGAGTTGA TAAAAACGGTACGGCAATCAGTCTGGAAGAAAACGGTTAGAACCAAAGAGTAATTACGCCTTACAGGTC TGTACTTTATGATAACGACGTGGTCAGATGGCAAAAACCTGAAAGCCGCTGCACTGGTGAGTTAGAA ATTACAGATATTAAACGTATTATCTTGAGCAGGGACGCTGCTGCGCATGATGGGGCGTGGCTACGC GTGGCTGGACACGGGACTCATCAGAGTCTGATAGAAGCAAGTAATTATTGCGACAATTGAAGAGCGCC AGGGATTGAAGGTTCTGCTGAAGAGATTGCAATTGCTAAAGGTTTATTGATGTTGAGCAAGTAAGA AAATTAGCTGACCTAAAGAATAATTGAGCAGTATCTTATAAAATGACGAAGGATTCAAATTA ATGAATGTGATTAGAACTGAAAGATGTCATAATTCTGGAGCCAAGAGTATTGGTGTGATGAGG TTTCTTTATGAGAGCTTAATCAATCAGCATTGAACATATTCTAGGCTATCCGGTCACTGGTCAAG ACAATCACTCACGTTCAAAAAATGTAACAGAGGCCCTCACTTCAACGCGGGCAGTACGCACAAGAT

AAACTTGTACGCTGCACTCATGGAGCAGTTTGATGTTGCTGTTGATATTGACCCAATTGGTATCCTT
TGGTAAATGGGTGGTCTGCTTCAGCTGATAATAAGCAGCAGTGTGGATACCAAAAGGGTTGCTC
ATGGCTTTGGTCTGCTGATATCGCTGAATTCAATATAAAAACATAAAACTATTATCATCCTGAAAGC
GATTGTGGAATATGTGGAATGATGAACGCATGCAATTGATTGGCCCCAACATCAGGGTTAATCCTTC
GCCAAAAGATGAAAGGCTTTACGTTAGATGAGCTTACGATTAATTAAATTGATGAGGCCGCGCTT
AAGGAGGACTAGTCCGGCGGCCATGAGTTAATCAAAACAGTTTGGAACCTTGCGGGTATGTACT
TCCAGCTATTGTGACACTACCAGCTTGGTATTATGGGCGAAATTAGGCCAGAATTATTGGTGTAT
TCACCTGGCATTAGCTGTTGGTTATGCAAGCATTGATGCAAGGCCACTCGCGCAGTGATACGA
GAAGTCGAATTGAAAAAGATAATGAAGAAAATAAGTTGAAAATTATTCTCAGCAGCTGTAATTAT
TTATTGAGTTGGCGCCTCACTTTATTATTTTTAGTGGTCATATCGCATTGCTACTGAACATTA
GTGAGACTTTTCATAATGTAAGTGTCTCGCTTAAATTCTCGCAGCATCCATACCATTATTTGATT
ACTCAAATATGGTGTCAATTAGAAGGTGAAGAAAGATTGGTTACTTAATATCTACAAATCAATTAC
GGGAGTGTATTAGCAATCTCACCGCATTATTATACTTAAACCCCTTTGATGTATGCGATAATAG
GCTTAGTTCTAGCAAGGTTTATGTTTATTGCTTATAATTGTCACGATAAAGTGCTTAAAGCT
AAACTAACATCGATATACCAACAATTAAAGATTGTTATGTCGGTGGATTACAGTAAGTAATAT
CATCAGCCCTGTGCTATCATATTGATAGGTTATTGTTCAAATCAACTGGGCTGCTAATGTTGCTT
TTTACTGCACCATCAGAAATTATTCTCGGCTTAGTATAATTCCAGGTGCGTTCAAGAGCCTTATT
CCAAGATTAGCTAATGCAAATAATTCCGCTGAAAGATATAAAAGAAAAGATTAAATTACAATTTCAC
ATAATCATCACCCCTATTGTTGATTGGCGTGTATTTCAGAGAAGATAATGGTTTATGGATGGGG
CATCATTGGTGAGCCTGGTTGGTATTCAATATTACTGATTGGCTTATTGATTGAGCT
CAAGTACCATTGCCAGTATTCAATCCCAGGTCTAGCTAAAGATAACTGCATTGTTCATCTTAGAGTT
GTTCCCTATTACTTTACCTCATAAAAGCACATGGGGTGTGGCGCGGTATTGCGTGGT
CAGTGGAGATGAGTAGATTATAGCATTAGTCTTGGACGGTAAAGTATATTAAATAAAATTCA
AAATGCAAGTTAAACTCATGGCTTATTGGTAGGTGACAATTATAATGATATATAATTAAACTTTA
ACTCTCTCTAGTTAGCCATAATGTTCTCTCGGACAAAAAGTAGGATCACATCTCATTAC
TTTGATTTTACCATGGTTACTAATTGTCGGGATAAGTAATTACGATCAATTGAGTTA
ATGAAAGAAGCTTTACTCTTGTGATTGGTTACAGTTATTGTTATTTTATTTGATAGGGAACTG
GTTAATTATAACGTGAAAATAATGTTATTGTTCTCACATTAAATAATGAAATGAAAAATA
TTGGATCATTGTCATCCCAATTCTATTACCATTTGAAATAATATGGTGTATGGGGAGCAG
ATGGATTCTTCTCAATTACGTCTGCAAATACATTGGAGGGCTATACGGGAAAAATTATCTTAAATG
CCTGCTGTATATCCTCTAAATGATGGCTATGTCGAATTGTTGCTAACAAAACCTCCAAATTAAATAA
ATACTCCATTATTCTGGATGTTTGATTTGATGGCACAATGGAAAATTCAATTAAACGCCAA
TATTGACATATTAAATTATTGACTTAAACATAGATAAAAGAAAAACAAATAAGTTACATTG
TTGATAATTATATTAGCTTAACCTTGCAATTACAGTATGGCTGAGAATGACCACTAACATTAC
TATTGAGGCTCTATTACCAATAATTGCTTGGCCAGTTGAATGAAGTAATAGTAGTCATT
TTGGTGAGTACGTTAGATTCTATATGCTATAACTAATAAAATTGCGCTTATTAAAGAATTGCCAGTA
AATACTATTCTGACTATTACGTTCTGTACCAACAAATGTATATACTGCACTTCACCAATTACCA
GGATTGGTTACTGGCATCATATTGGAGCAGTATTACGGACTAATATATGTGAGTTATCACCG
CCGGTGTGTTGGAAATAATACACAGGCATTACTGATTACGCTTACGATTGTTGAGTACGAGCCT
TTCTGCTGAAACGCTAGTAACGAAATTAGCTGAAATGTGATGTTAGTATTGACCATCTTACTATG
GCGATTACAGTAATATGCAAACACAGTACAGTAACCTTAATGCCACCTACAATGGCGAGGCCT
CAAAATCAGATTGTCACTACAACAAACATTCTAACTGGCGTTATTTCAGGATGATGGGT
CTACAGACAATACTATATAATAAAACTCCAAAATCTGACTCCAGAATTGGCTAGTTGATGAT
AATTGAAAGGTCAAGGTGCAAGGAAAAATTGCTGATAAAGTACAGCGAGACAGATTACAAAT
TTATTGACCAAGATGATATTGGTAGAAAACAAATTGAAATTAGTAAAGTATGCAAATGAAATT
AATTGAATGATCAGATGCGCTTCGCTAGTTATGCTGATGGCTATGCTTATATGGATGGTGGAGGTACA
ATCGATTTCGGATATCTAACATCATGCTGATCAATTAAAGGATTGTTTTAATGGTGGATA
CCAAGGATGTTCTATTGTCATCGCAATGACCAATTCTCTGAATTATCGAGGATTGATATC
TACATGACGATATCACAAACATTAGCTGCACTACGCTTGGTAAAGTTATTCTCCGAAATACCTTATG
TTATATAGACAGCACACGAATGCGGTAACGGTATCAAACATTCCGCAATGGATTGACTCTAAATTAA
ATCACCACTAAACTATCTTATCACGAAAACATTACGTTAAAGGAAATTCTTGAATGTAACAGCT
CTATCTTATCAGAGACGAATAAAAGTTGGATTATTGATTTGATGAACTAAATAATTAAATT
ACAGATTGTTAAGTTAGCTGAGGTGGTTAGATTAAATAACAGTAGAAACTAAATTATTAAATT
CTTAATACGGAGAAAATTAGCGAATGATTCAACTACACCTACTTAAATCGGCAACATACTTATC
AAGGCTATTCAATTCTTATTAACAAACTGATAAAGGATTGAGTGGATAATAATTGATGATGGTAGTA

TAGATGCAACAGCGGTACTTGTAGAAGATTAGAAAAAAATGTGATTTGACTTGATTATTGCTATCAG
GAAAATAATGGTAAGCCATGGCTTAAACGCTGGTAAAGCTTGTAGAGGCATTATCTTATTGT
TGACAGTGATGATGCACTAACCTCCGATGCCATAAAATTAAATTAAAGAATCAATACATGATTGCTTATCTG
AGAAGGAAAGTTTCAGCGAGTCGGTTAGAAAAGCATATAAAAAGGGGGATTATTGGTAATGATT
ATAATTCTCAGAACATATATACTATTAAATGCGACTGAGATTAGCAATTAAATGGTATGTTGC
ATATTGTTAAAAAGAAAGTTGGTAAAAATCCATTCCCCGTATAAGAGATGAAAATTGTTCCAG
AATTATATATTGGAATAAAATACTGACAAGCGAAGATTGATTAAACATAAGCAAAGTTATATCTT
TGTGAGTATCTGATGATGGCTTTCTAAAATTCCATAACCAGCTAAAAAATACCCAAGGGTTAA
GATTATTACAAAGATCAAAGAAAAGAGAGAAAACCTATATAAAAAAAACAAAGATGCTAATTAGATATT
TGCAATGTTTATTATGAGAAAATAAAACTATTGTCATTACAGGTTAGGCCTGGAGGTG
CTGAGAAGCAGGTTGCTTAGCTGATAAAATTAGTTAAGCGGGCACCATGTAAGGATTATTCACTT
GGACATATGTCTAATAATAAAAGTCTTCCTAGCGAAAATAATGTTAATGCTTAAATGTAATGTCAAA
AAACATTCTGGAGTTATAAAAGGTTGTCTAGAATTAGAGATGTTAGCTAATTCAAACCGACATTG
TACACAGTCATATGTTCATGCAAACATTACTAGATTGCTGTAATTGGAATCAAAACAGACCTGGT
ATTATATCAACTGCACATAATAAAATGAAGGTTGGTATTTCAGAATGCTCACATATAGATAACCGATTG
TTAAGTGATTGTTGACAAAATGTTAGCAAAGAAGCAGTGGATGAGTTTACGGATAAAAGCCTTAATC
CCGCTAAAGCAATTACTATGTATAATGGGATAGATACCAATAATTAAATTGATTATTGGCAAGGAGG
GAAATTGAGACGGTATTATAATAAAATGATGATATAATTACTTGCTGCAAGGCGTTAACGTTAGC
TAAAGATTATCCTAATTATTGAATGCAATGACTCTGCTCCTGAAACACTTTAAACTTATTATTGGTG
ATGGTAATTGCGTGACGAAATTAAATGCTTATAAAAATTGCAATTATCTAATAGGGTGTCTTGTG
GGAGTTAAAAAAATATTGCTCCATTTCATGATGTGATATTGTTCTCTCTCGTTGGAGG
ATTGGATTAGTCGTTGGCAGAGCTATGTCATGTGAGCGAATTGTTGGCACGGATTAGGGGGAGTAA
GAGAAGTTATTGGTGACGATGATTCTTGACCGATCTGATCACATTGGTTTCGGAATCGTAGCGTATT
AAATTGCTTTGAGCCAGATACGTGATCACATTGGTTTCGGAATCGTAGCGTATTAAAAATTCTC
AATAGATACTATTATTGCACTGGCAAGAACCTATGAACTATAATTGCTCAAAACATGAAAGGTAGA
TTTATATTGGAACGTGTTGAAATTAAATTCAATCTCAATTGAGATTGTTGATTCAAAAATA
CCATCATAGCTAACGATGATTGGTATTATTAAAGATGCTTCTATAAATATATTGACGTTTAAATCG
CCGAAACGATTGGCTGGAACAGAGAAGTAAACCTGTTGAGAATGAGATTGAGATGTTATGG
ATATTAAAATTGATCCAGTGAATTAAATTATAATAAAATCAAGATTAAATGTTAATAATGATAATCT
TTCTGACACTCATTAATTATGAGTGGTACGTTGGTAAACGGTAAACTATTATGACAGCTAGAAC
ACTAAAGTTTGCACTTACAATTACTCCCACCTCTAAGTGGCGTCAAAGGGTAACATTAAACGAAATTAG
TGCCTTATATACTGATTATGATTACACTAGTTGCTAAAAAAAGGTCCACTAACAAAGCATTGCTGG
AATATGATGTCATTGTCATTGATCCCCGAACCTACGAGAGAAATTACCGTAAGAATGATTAAAGCA
TTGTTCAAGCTTATAAGTCATAAAAAGAAAATTGACATTGTCATACACATTCTCAAAACAGG
TATTGTTGGGCGAGTTGCTGCCAAATTAGCAGTGTGAGGATCCACACTGTACATGGTTTCTT
TTCCAGCCGCATCTAGTAAAAAAAGTTATTACCTTATTGTTCATGGAATGGATAGCAAAGTTCTTACG
GATAAGTTAATCGTCTGAAATGAGATGATGAAATATAGCAATAAACAAATTAAACAGGGATAA
AGTTTTTAATTCTAAATGGAGTAGACACTGATAAGTTCTCCTTAGAAAATAAATTATAGTAGCA
CCTGAATCTAGTAATGGTGGTAGATTATCCAAGCAAAAGATCTGAGACATTGCTGCTGTTGAA
AAACTGCTGAATGAAAATGTTAATGTTAAGCTGACACTTGTAGGAGATGGTAACAAAAAGAACAGTTAGA
AAGCAGGTTCAACGGCAAGATGGACGTATAATTTCATGGATGGTCAGATAACATTGTTAATTAA
AAAGTTAATGATTTTATTACCTTCTTGGGAGGGTATGCCATTAGCAATTAGCAAGCATTGAGC
TGTGGACTCCATGTATAGTCACTAATATTCCAGGTAAATAAGCTTAATAGAAGATGGCTATAATTGTTG
TTTGTGAAATTAGAGATTGTCAGTTATTACTCAAAAATCATGTCATATTGTTGTAAGCCAGAACGTA
TTGCACAGCAATCTACCAATGCACGATCATTTATTCTGAAAATTATGGATTAGTTAAAGAAATAAAG
GTCAGACAGCTATATGATAATTAAAGAGCTCGGTACCCGGGCTAGGGTGTAGGCTGGAGCTGCTCGAAGT
TCCTATACTTTCTAGAGAATAGGAACCTCGGAATAGGAACATAAGGAGGATATTCAATCCGTCACGGCG
CCGCCCTGCAGGCATGCAAGCTGATCCATGGATCGCTAGCTTAATTAAATAAGCCGTAAGCATATAA
GCATGGATAAGCTATTATACTTTAAATAAGTACTTGTATACCTATTGCGAACATTCCAGGCCGAGCA
TTCAGCGCGGTGATCACACCTGACAGGAGTATGTAATGTCAGCAACAGATCGGCGTAGTCGGTATGGCA
GTGATGGGACGCAACCTTGCCTCACATCGAAAGCGTGGTTAACCGTCTATTGTTAACCGTCCCG
TGAGAAGACGGAAGAAGTATTGCCAAAATCAGGCAAGAAACTGGTCTTACTATACGGTAAAGAGT
TTGTCGAATCTGGAAACGCCCTCGCATCTGTTAATGGTAAAGCAGGTGAGGCCACGGATGCTGCT
ATTGATTCCCTCAAACCATATCTGATAAAAGGAGACATCATCATTGATGGTGGTAACACCTTCTCAGGA
CACTATTGTCGTAATCGTGAGCTTCAGCAGAGGGCTTAACTTCATCGGTACCGGTGTTCTGGCGGTG

AAGAGGGGGCGCTGAAAGGTCTTCTATTATGCCTGGTGGCCAGAAAGAAGCCTATGAATTGGTAGCACCG
 ATCCTGACCAAAATCGCCCGTAGCTGAAGACGGTGACCATGCGTTACCTATATTGGTGCGATGGCG
 AGGTCACTATGTGAAGATGGTCACACGGTATTGAATACGGCGATATGCAGCTGATTGCTGAAGCCTATT
 CTCTGCTTAAAGGTGGCCTGAACCTCACCAACGAAGAACACTGGCGAGACCTTACCGAGTGGAAATAACGGT
 GAACTGAGCAGTTACCTGATCGACATCACCAAAAGATATCTCACCAAAAAAGATGAAGACGGTAACCTACCT
 GGTTGATGTGATCCTGGATGAAGCGGCTAACAAAGGTACGGTAAATGGACCAGGCCAGAGCGCGCTGGATC
 TCGCGAACCGCTGTCGCTGATTACCGAGTCTGTTGCACGTTATATCTCTCTGAAAGATCAGCGT
 GTTGGCGCATCTAAAGTTCTCTGGTCCGCAAGCACAGCAGCAGGCGACAAGGCTGAGTTCATCGAAA
 AGTTCGTCGTGCGCTGTATCTGGCAAATCGTTCTTACGCCAGGGCTTCTCAGCTGCGTGC
 CTGAAGAGTACAACACTGGGATCTGAACATACGGCGAAATCGCGAAGATTTCGTCGCTGCATCATCCGT
 GCGCAGTTCTGCAAAAATCACCAGATGCTTATGCCGAAAATCCACAGATCGCTAACCTGTTGCTGGCTCC
 GTACTTCAAGCAAATTGCGCATGACTACCAGCAGGCGTGCCTGATGCTGTTATGCAGTACAGAACG
 GTATTCCGGTTCCGACCTCTCCGCAAGCGGTTGCCTATTACGACAGCTACCGTGCCTGTTCTGCCGCG
 AACCTGATCCAGGCACAGCGTACTATTGGTGCCTACTTATAAGCTATTGATAAAGAAGGTGTGTT
 CCATACCGAATGGCTGGATTAA

SEQ ID NO: 18 (example O25B rfb locus nucleotide sequence – O25B-EPA production strain
stGVXN4459)

ATGACGAATTAAAAGCAGTTATTCCTGTAGCGGGCTCAGGATGCATATGTTGCCTGCCACTAAGGCAT
 ACCCAAAGAGATGCTACCAATCGTCGACAAGCCAATGATTCACTGTTGACGAGATTGTGGCTGCAG
 GGATCAAAGAAATCCTCTGGTAACTCACCGTCCAAGAACGCGGTCGAAAACCACCTCGACACCTCTTAT
 GAGTTAGAATCACTCCTTGAGCAGCGCTGAAGCGTCAACTGCTGGCGGAAGTACAGTCCATCTGTCGCC
 GGGCGTGAACATTATGAACGTGCGTCAGGGCGAACCTTAAAGGTTAGGCCACTCCATTGTGTCGCGAC
 CTGCCATTGGTACAACCCATTGTCGTTACTGCCAGACGTTGTGATCGACGATGCCAGCGCAGCC
 CTACGTTACAACCTTGTGCGCATGATTGACGTTCAACGAAACGGGCCGAGCCAGGTGCTGGCAAAACG
 TATGCCGGGTGACCTCTGAAATACTCCGTACCCAGACTAAAGAGCCGCTGGACCGTGAGGGTAAAGTC
 GCCGCATTGTTGAATTATCGAAAACCGGATCAGCGCAGACGCTGGACTCAGACATCATGGCGTAGGT
 CGCTATGTGTTCTGCCGATATTGGCGGAACCTGAACTGGAACGTAACGCTGCGATGGGACGTATTCA
 GCTGACTGATGCTATTGCGAGCTGGCAAAACAAATCCGTTGATGCAATGCTGATGACCGGGCAGAGTT
 ACGACTGCGGAAAAAAATGGCTATATGCAGCGTTGATGAAAGTATGCCCTACGCAACCTGAAAGAAGGG
 GCGAAGTCCGTAAAGGTATTGAGAAGCTGTTAAGCGAAATAATGAAAATCTGACCGGATGTAACGGTTGAT
 AAGAAAATTATAACGGCAGTGAAAATTGCGAGCAAAAGTAATTGTCGGAATTCTCCTGCCGTTTTTA
 TATAAACCATCAGAATAACAACGAGTTAGCAGTAGGGTTTATTCAAAGTTCCAGGATTTCCTGTT
 CCAGAGCGGATTGTAAGACAATTAGCGTTGAATTTCGGGTTAGCGCAGTGTTAACGCTCGTCAC
 ATCATAGGCATGCAGTGCAGTGTGCTGGTAGCTGAAAGCCAGGGCGGTAGCGTGCATTAAACCTCTATT
 AATCAAACGTAGAGGCCCTTATTACAGCATGCTCTGAAGTAATATGGAATAAAATTAGCTAGCAGTGA
 GATACTGTTACTGGGGCGCAGGATTATTGGTCTGCTGTTGTCACATAATAAAATAACGCAAG
 ATAGTGTGTTAATGTCGATAAAATTAAACATACGCCGAAACCTGGAATCACTGAGATGTTCTGATTCT
 GAACGCTATTCTTGAAACATGCGGATATTGTGATGCACTGCAATGGCACGGATTGGCTCAGCATCA
 GCCGGATGCAGTGATGCACCTGGCAGCTGAAAGCCATGTCACCGTCAATTACAGGCCCTGCCATT
 TTGAAACCAATTGTTGACTTATGTCCTTTAGAAGCGCTGGAATTATTGGTCTGGTCTGGATGAT
 GAAAAGAAAAAAACTCCGTTTCATCATATTCTACTGATGAGGTGATGGTACTTACCCATCCGGA
 TGAAGTAAATAGCAATGAAACGTTGCCGCTATTACGGAAACGACAGCATCGGCCAAGTAGTCCATT
 CTGCTTCTAAAGCTTCCAGCGATCATTGGTTCGCGCATGAAACGTTACTTATGGTTACCGACCATG
 ACTAATTGCTCGAACAACTATGGTCCTTATCATTCCGAAAAGCTTATTCCACTGGTTATTCTAATT
 ACTGGAAGGTAAGGCATTACCTATTATGGCAAAGGAGATCAGATCCGCACTGGTTGATGTAGAGGATC
 ATGCTCGAGCGTTATATAACCGTCGTAACCGAAGGTAACGGGGCGAAAATTATAACATTGGTGGACACAAC
 GAAAAGAAAAACATCGACGTAGTGTCACTATTGTGATTGTTGAGATGAGATAGTCCCAGAGAGAAATC
 TTACCGCGAGCAAATTACTTATGTTACCGATCGTCCGGGACACGATCGCCGTTATGCGATTGATG
 AGATTGGTGCAGAATTGGGATGGAACACAGGAAACGTTGAGAGTGGGATTGCTAAAACGGTGGAAATGG
 TACCTGTCCAATACAAAATGGGTTGATAATGTGAAAAGGGTGCCTATCAATCGTGGATTGAACAGAACTA

TGAGGGCCGCCAGTAATGAATATCCTCCTTTGGCAAAACAGGGCAGGTAGGTTGGGAACTACAGCGTGC
TCTGGCACCTCTGGTAATTGATTGCTCTTGATGTTCACTCCACTGATTACTGTGGTGAATTAGTAATC
CTGAAGGTGTAGCTGAAACCGTAAGAACGATTGGCCTGATATTATGTCAACGCAGCCGCTCACACCGCA
GTAGACAAAGCAGAATCAGAACCGAAGTTGCACAATTACTGAACCGCAGGTGCGAACCGATCGCGAA
AGCAGCCAATGAAGTCGGCGCTGGTTATTCACTACTCTACTGACTACGTATTCCGGGGACCGGTGAAA
TACCATGGCAGGAGGAGGATGCAACCGACCCTAAATGTTACGGTAAACCAAGTTAGCGGGAGAAAAAA
GCATTACAAGAGCATTGTCGAAGCACCTTATTTCGGGACAGCTGGTCTATGCAAGTAAAGGAAATAA
CTTCGCCAAAACAATGTTGCGTCTGGCAAAGAGCGTGAAGAATTAGCCTTATTATGATCAGTTGGT
CGCCAACGGCGCAGAGTTACTGGCTGATTGTACGGCACATGCTATTGCGTGGCACTGAATAACCGGAA
GTCGAGGCTTGTACCATGGTAGCTAGTGGTACCAACGTGGCACGATTATGCTGCGTGGTTTG
AGAGGCAGCAGGAAAGCAGGCATTCCCCTGCACCTACAACAGCTCAACGCAGTACCAACACAGCCTATCCTA
CACCAAGCTCGTCGTCACATAACTCTCGCCTTAATACAGAAAAATTTCAGCAGAACCTTGCCTGCTT
CCTGACTGGCAGGTTGGCGTGAACGAATGCTAACGAATTATTCAGACTACAGCAATTAAAGT
GCATCTGTTGTAATGGTGGAGCAAGATGTTAAAAGGAATGATGAAATGAAAACCGTAAAGGTATTA
TTTGGCGGGTGGTCTGGTACTCGTCTTATCTGTGACGATGGCGTCACTAAACAGCTGTTACCGATT
TATGATAAAACCGATGATCTATTACCCGCTCTACACTGATGTTAGCGGGTATTGCGATATTCTGATT
CAGTACACCACAGGATACTCCTCGTTCAACAACTGCTGGTGACGGGAGCCAGTGGGCTGAATCTC
AGTACAAAGTCAACCGAGTCGGATGGCTTGCAGGGCGTTTATTACGGTGAAGAGTTATTGGTGGT
GATGATTGTCGGTACTGGTAGATAATATCTTACGGCCACGACCTGCCAGTTAATGGACGTTAGC
TGTAAACAAAGAAAGTGGTCAACGGTATTGCTATCACGTTAATGATCCTGAACGTTATGGTGTG
AGTTGATAATAACGGTACTGCAATTAGCCTGGAAGAAAAACCGTGGACCACAAAAGTAACATGCGTT
ACTGGGCTTATTCTATGACAATGACGTTGGAAATGGCGAAAAACCTTAAGCCTCTGCCAGGTGA
ACTGGAAATTACCGATATAACCGTATTATGGAACAAGGACGTTGCTGCGTATGATGGGCGTG
GCTATGCGATGGCTGGATACAGGGACGCATCAAAGTCTTATGAAAGCAAGCAACTCTATTGCCACATTGAA
GAGGCCAGGGACTAAAGGTTCTGGGAAGAAATTGTTATCGTAAAGGGTTATTGATGCTGAGCA
GGTAAAAGTATTAGCCGAACCGTTGAAGAAAAATGTTATGGTCAGTATGCTCAAATGATTAAAGGTT
ATTAATAAGATGAAACGTATTAAACTGAAATTCTGATGTCGTTGAACCAAAAGTTTGGGA
TGAACGTTGGCTCTTTGAGAGTTTAATCAGAGGATTGAAAGAAGCAGTAGGTCGTAAGGTTGAGT
TTGTCAGGATAACCATTCAAGTCAGTAAAGGTGTTACGTTGCTTCTCATATTAGTAAACCTTAT
GCTCAAGGAAACTGGTGCCTGTTGGCGAGGTTTGATGTTGCGGTTGATATTGCTAAATCGTC
ACCTACATTGGGAAATGGGTTGGGGTGAATTGCTGCTGAGAATAAGCGTCAGTGTGGATTCTGAGG
GATTGACATGGTTTGGTGCTGAGTGATTAGCAGAAGTTATATAAAACGAATCAATTATGCT
CCATCACATGAAAAAAATTATGGAATGACCTCTGCTTAATATTAAATGGCGAGCACAGCACTGAT
CACTCTGCTGATAAGGATGCAAATGGGAAAGATTGAAACTAAGTGAGTTGAAATGTCCTCTAAAA
CATAGTATATGGAATGTTGGGCTACTTTATACCAACATTAATTGCAATTCCGCCTTGGATTAATTG
GAGGAAAATTGGTGTAGAATTGTTGTTGATACGTTAGCAATTAGTTATAGGGTATGCAAGTATAT
TTGATGCTGGGTTAACAAAGAGCTGTTGCGTGAATAGCATTACTAAAAACAGAGTGGACGATTGTAAT
ACGATAATAGTAACCTCTATTATCGCTGATATTGTTAGGGTTATCGGAGGGGGAGTGTGTTCTGCT
TAAAGGCATATTGAACTGTTAAATATCTACCAATATTACGCGATTGATGAAAGTCTCTAGTAT
TATTATCATCTGATACCTGATTCTAGTCAGCAGGAAACTATTAGCAGAGCTTGAGGGTCGGGAAATAT
TTTGGGATTCTAAATATACAAAAAAAGTGTAGGAAATTCTTAATTGCAAGGGTTACCTGCATTATTGTTT
AATTAAACGCTTTCTGCAATTATTGGTAGGCAAGAGTTATGCTTGTGGTAAGCT
ACATTATGAGCAGGGAAAGAATAACTATCGATATCTCATTTTCAATAACTGTTAAAGCGGTATTT
AGATATGGCGGGTGGTAACTATAAGTAACATAATCTCTATATTAGCAGAGTATGGATAGATTATTCT
ATCCCATATCCAGGGAGCATCAAAATATCATTCTACAGTCCCTAATGAGCTGGTAACAGGCTTGGAA
TAGTCCAGGCTCTTGGGAAAGCTGTTTCCAAAATAAGTCATGCAAGGAATTACAGCGTCATAT
GCAGAGCAAAAAAAAGCTTATATAATGACTGTCATTGTAATGCCATTGGTTATTGATATT
CGAAAGTTATTAAACATTGGGATGGGGCTGAGTATGCAAGGGATTGCGTCAATATTACGGATTA
TGCTTATAGGGTATATTAACTGTTATTCAAAACTCTTTGCAACATACAGGCCATTGGAAAAGCA
AAATACACTGCATACATCCATATGAGGAAATTGTTATTGATAATGTTATATATAATTCAAAGGA
ATATGGGTTATTGGTAGGCGTGGTATGGACAATTGAGTAATAATTGATTGTTATTAAATTGAA
TGAGTTATGTTGTAATAATCTTATGAAAAAAAGGGTAGCCTGATGATATATTGTTATTAAATTGAA
TGGGGCTATAGATACCATTAATTGTTAAAGTTAATGGATTAAATGTTAGCGATTATAAAATTATCA
TTGTTGATAACTGTTCTATGGATAACTCATATGATACTATAAAAGAAAATCTTAAATTATATTGCT
GATAAAAGTATCATTGAGGTGAAGTATGAGGATAGAAATAAAACCTTAAAGAAAACGATAAAATCAT

ATTAATACAATCTCGCAAATAATGGGTACGCAAGTGGTAATAATATTGGCATAGAGTCGCTCTTAATC
AGGAGAATATGAAATACTGCTGGGTCTGAATAATGATACTGAAGTGGATAAAGAGGCTTAACTCATT
ATTAGTAAATGTGATTGAGATAAAAAGTATAGGGATTGCGGTTCTCGTTAGTCTATTTGCCGACAGAGA
GATGCAGCAAGGACTAGGTGGGGTGATAACAAATGGTTATGCACTACAAAAAATTATGAAATGGGAGAT
TAGTTCCAAAAAATATGATGATGAAGTCATTGATAATGATAATAGATTATATAATTGGCGCATCGATGTT
TTCTCTAGAGAATGTTGAAACAGTTGGATTGATGAATGAAGAATATTTTATACTATGAAAGAGTTAGA
TATTGCCTCAGAGCAAAGCAAAGAACCTTAAATTAGGTATTGCTCAGAAAGTTGGTTATCATAAAA
TAGGTGCAAGTACTGATGGGGAAAGAGCATGATGGCTGATCTTGCCTAATAAAAATAGGCTGGTCAATT
ACAGAAAGGTTTATCCCATAATTGAGCGGTATGGTGTCACTTTGTTGAGCATTAAACCGTGC
TAGAAGAGGTGAGTTAATAAGATGAAAAGATGTTGAATGTTATGTTAACTCAAACGAAACAAAGGTA
GCAAATGCCATTAGAATAATGCACTTAATCATGGTTAATAATCTATAGTTGATATGTTATTAAAGGGT
ATTTAATGAAAGTGGCTTTTATCTGCTTATGATCCACTATCTACATCCAGTTGGCTGGCACACCTTAT
TATATGCTAAAGGCATTATCGAAGAGAAATATTCCATTGAAATATTAGGACCGGTAATAGCTATATGAT
ATACATGTTAAAGTATATAATTAAATTAAAGGTGTTGGAAAAGAATATGATTATAGTCATTGAAAGT
TGCTTCCAGGTATTACGGTAGAATATTGGTAGGAAATTAAAAAAATTGATGGTTGGATTATTATC
GCACCTGCAGGGTCTCACAATTGCTTTAAAACACCATAACAAATAATATCTATCGGATACAC
ATATGATCAATTAAAAGCTATTATCGAATTAAATAAAAACAATTATAATGATGAGGATGCAAGTT
TAATCGAACGCAAGGCTATTGAAAAGCAACAGTAGTATCTTCCATCTAAATGGCAATGGATTG
AGGAATTATTACAGATTAGATTGATAAAATTAGTTGAAATACCATGGGGGCTAATTATTGATGAT
TCACTTGCTAATAAAATATAATTCAAAAGAATAGTTACCTGTCTTTCTGGAGTTGATGGAA
GAAAAGGTGGAAAACAGCCTTGAAAGCAATTGAAATATGTAAGGCAATTGTTATGGGATCGATGTTAGACTA
AAAATTGTGGATGACTCCGAATCAAAAGATTACCTACTTGGGTGATTAAATTGATAAAAGTAGATAA
AAATAACGTTGACGAATATCAGAAATTCTCGATGTGTTATCTAACGCTGATATACTTCTTACCAACCA
TTGCTGAATGTTATGGAATGGTATTGTAAGCTGCTGGATTGCTGTTGCGCTACAGATA
GGTGGAGTCAGTTCTATAGTTATCAACGAAAGGACGGGGATATTAAATTAAAGACCCGTTAGACTATAAGCA
CTTGGAAATGCAATTCAAAATAATTAGTCCGTAGAGACTTACAAAACTACTCCAAAACGCAAGAA
TTAGATATAATAATTGCAATTGGGACAATTGGCTAAAAGATAATTGAGATTATGAGCATAAG
AATAGAAGAATCAAATAGCACAAAAGAATTATGTTATTATCTTCTTCTGTTCTGATT
TGTTTATACATTAGGGGTGATAATTAGCATTCAACGATAATCTCAATTACATTGCTTTGTT
TTAAGAGCTAAAATATTGCAAGATAATTCTAATAATAGTAGCGTTATTCATATTGTTGTT
CTGTTGTTAAGTATGCTATTAAATTGAAACAGGCTTAACATTAAAGTTGACTTTCAATATAAGCA
CTTAAATATGGCATACTGCTCCTCTGTTATGCACAGACGTTGTTATGTTCTGAAAGAAACTTAAG
AGATCCGTCTTATTGTCGATTCTTGCCTATTGGCATTATAAGTATTCTTACAGAAGACTGA
GATTATACATGATAAAAGTATGATTCTTCTGAAACCCTACGCTTGCATTGGTTTACACCTATCT
TTCTTCTGTTATACATACAAGAGGGGGGGCTACTATTGCTCTATATATTCTTGGGTATTGCG
TTAGGTATCCAGAATTAAACAATGTTGGTAGGCATTGTGATTAGTGTGTTGATGAAAAAAATAACTAT
AAGGCAAACATTGTTACTTTGGGGCATGGATTCTTCCATGATATTAGTATTAGACATTCTT
ACTATACATCGGGCTTGATTTAAAATACTACGAACCTATCAGTGCCTGTATATCTTCAGGAATTGAA
AGAGCTTCTGAAATTATTACAAGTTATGGCTTGGTATTGGTTTCAACAAATGGAGTGAATGGGA
GATAGGAATATCAACAAATTAGCTGAACCTTGATGCCCTATGTTAAATATAACGATGGCTCATT
TTCTCTAAGTTAATATCTGAGTTGGGTATTGGTCATTAAATGTGTTATTCTATTGTT
TCCGATTCTGCTTCAAAAAAGTAAGAGATATTACCGCAGTATATTGATATAGCTTCTA
CATGTGTTCTCATCCCTTTTATACGTGGTGTGGTTATATAACCCCTATGTTATTGTT
CATCAATATTGTCGAATATCACGCTAAAATCTGATGAAATCTAATGTCCAGATAGCTATATAA
TAGTAGATTATATTACATTACGTAATTACATATTAGCATATGATAACTAGGACATAATAA
TGTGCATTAAAAAAACTTAAGTTAATTAAACGATATGCCATTGGTGGTCTAGGCTTAAAGAT
ATATTCTTAACAAATTGTTATTGTCATTGTTAGGATTATTAGATTCCATGTTATTAGAAAAGA
TGGAAGTGTAGTTGGAAAAGGTTTACATCAGGTGAGGATTACGAGTTGATGCATTATGGATGCG
TAGTTCCATTGGAGAAAATGTCATTAAATGACTATGTCACATCGGGCTATTAAATATGTCATT
GGTAGAGATACTTAATAGCAAGTAAAGTATTGATGATCATAATCATGGTATTGTTCTAAATCCGA
TATCCATAGTTCACCAACTATTATCCTCGCTAGGCCCTGAACTGCACTGTGTTATTGGAGAGC
GTGTGTTGGATTGGCGAAAATGTGACAATTACCAAGGTGCGTGTAGGTAATTGGTAGTTATTGCG
AACAGTGTGTTGCGTGGTGGAGATTCTAATAATGTGATCATTGCTGGTCTAGCTAAATGTT
ATATAACTATGAGCGTATGCAATGGGAAAGAATATAGTTGAAATATCGGCTGTTAATT
CCCCCTTACCGTACTAAAAATGTGCTTACAGCAACTAAAGATAGAGGCCAATGTAATT
ATTGCACTGG

TTCATAGCTCTGCTGAACTAATGGAATTATTCGGTGGGTGAATTATAGAGTATCCAGAAGTCAGTCTCGTGGGTTAAAAGATTATTCGAATATAACTGCATAGATTCTAACGGTATTAAAGCAACTCATGGGTATGCTACATGATTACAGCAAATGTTAGTGACCCCTATAGATTGTTATTGCCACAATCCTGCAACGGTTCTATAAATATTAAAGCTATCGAGATATTAGGAGAACCTAAATTATCTTTTATCTTTTATGGGCTTTATACAATATCAATATAAAAAGAACACAGCAGTTTGTTCAGCAGCAGTGGCTAAAAAAAGAATTGCAAAAAAAATAAGTTAAAGAACATGGTGTAGTCGCCCTGAAGATATTGCCCTTGTAAAAGTATGGTTGTAAGAAATAATAAAAAAGGATGTGAGGATATTACCCAGCAGTGCCTCGTATAATTAAAGTCAACTTTGAAAGTATCATACGTGTCACAAATATTACAAGATAAAATATCATTATCTTACTTTTGATGGTACTGAAAATAAGTATGCAAAAGAACATATAATTAGCTCCGAACGTAAAATGTACATTCTCGGTTACCTTAATGCAACCGAGATGGTTAACCTTATCAAGATTCAAGATATTATTGTTCCCCTCGAAAATAGAAACGTGGGATTACCATTACAGAAGCTAAAACATACAAAAAATGGATATTGCCAGACTTACCTTATGCTCATGAAGTTTATATAACTATTCAAAACTAGATATTCCATTGACGATGAGAAAATACTTGTTCGCTACATATTAGAGTACACAAGTAAAATATGCATGAAGATATAAAAATAGTAGGGTGAATTAAATATGATGCATTGACTGGTTGAACAGTTATTGAATATATCCTCAAGGGAACTGACGTGGTTATATTAAATACGTTTCACATGGCCATGATGACTATATAGAAAATCTTTATTAAATTAAAGTGTGCCCTCTGGAAGATTAAAATAATAGTCGTGATAACAAAAGTCAATGGTTTAAACATGCGAAAAAAATTGCGTAACCTATTGCGATGGGGCAATATGGATTGGACATAATAAACATAGCAGTGTATATAATTAAACTTCATGATTATGAAATGATTATTCTCTTCTTAACCCGATGTATTCTAACCAAGTGAAGTTGATTAAAGCAACATGATTATTCAATACGGAGTTCCAACTTATATGATTCTTGTCTTTTATTGGGGTGAATAAAAAGTAAAATAAGAAGGAAATACTTCTGATACTGTAGTTGATTGGTGTGCTGGCTCATTTATGCTTATTCTGCTTAAGTTCTAAATGTGAATGGTTTGTCAAAATATTATGTATTGTGAAGATTGACCTTTGTATGCGTTAAAATTAAGTGGAGTAGACTTTACTATACTCCCCATTGATGCTATTCTATGCGCAGCATGAAAATAGAAGAATATTACTAAAGCATTGCGATGATGGCATATAAGGAGTATTACGCGCTACATATTACGAAACCAATTCTTCTTATAAAAACATAGAAAATTACATCCGAACGTGGTAAAGTGTGATTAAAGGATCCGTGAGGCTGCTCGAAGTTCCTATACCTTCTAGAGAATAGGAACCTCGGAATAGGAACTAAGGAGGATATTCTATGGATAAGCGTAAGCATATAAGCATGGATAAGCTATTATACTTTAATAAGTACTTTGTATATTGCGAACATTCCAGGCCGCGAGCATTGCGCGGTGATCACACCTGACAGGAGTATGTAATGTCCAAGCAACAGATCGCGTAGTCGGTATGGCAGTGTGATGGGACCGAACCTTGCCTCAACATCGAAAGCCGTGGTTACCGCTCTATTTCACCGTCCGAGAAGACGCTATGAATTGGTAGCACCAGTCCGACCAAAATGCCCGCGTAGCTGAAGACGGTGAACCTCGCTCGCAGTACCTGTTAATGGTGAAGCAGGTGCAGGCACGGATGCTGCTATTGATTCCCTCAAACCATATCTGATAAAAGGAGACATCATCATTGATGGTGGTAACACCTCTCCAGGACACTATTGCTCGTAATGTGAGCTTCAAGGGCTTAACTTCATCGGTACCGGTGTTCTGGCGGTGAAGAGGGGGCGCTGAAAGGTCCCTCTATTATGCCCTGGTGGCCAGAAAGAACGCTATGAATTGGTAGCACCAGTCCGACCAAAATGCCCGCGTAGCTGAAGACGGTGAACCATGCGTACCTGTTGACGCTTCAACATCGAAAGTACCGGTAAATGGACCAGCCAGAGCCGCTGGATCTGCCGAACCGCTGTCGCTGATTACCGAGTCTGTGTTGCACGTTATCTCTCTGAAAGATCAGCGTGTGCGCATCTAAAGTTCTCTGGTCCGACAGCCAGCAGGCCGACAAGGCTGAGTCATCGAAAAGTTCGTCGTGCGCTGTATCTGGCAAAATCGTTCTTACGCCAGGGCTTCTCTCAGCTGCGTGTGCGTCTGAAGAGTACAACACTGGGATCTGAACACTGGCAAATCGGAAGATTTCGTGCTGGCTGCATCATCCGTGCGCAGTTCTGCGTGTGAGAAATCACCAGTGCCTATGCGAAACGCGTGGTCAACATCCACAGCAGGCGCTGCGTGTGAGTCATCGAAAGTACAGAACGGTATTCCGGTCCGACCTCTCCGAGCGGTTGCTTACCGACAGCTACCGTGTGCTGCTGTGCGTGTGCGTACTATTGTTGCGCATACTTATAAGCGTATTGATAAAGAAGGTGTGTTCCATACCGAATGGCTGGATTAA

SEQ ID NO: 19 (example O75 *rbf* locus nucleotide sequence – O75-EPA production strain
stLMTB11737)

ATGACGAATTAAAAGCAGTTATTCTGTAGCGGGCTCGGGATGCATATGTTGCCACTAAGGCGAT
ACCCAAAGAGATGCTACCAATCGTCACAAGCCAATGATTGACGAGATTGGCTGCAG
GGATCAAAGAAATCCTCTGGTAACTCACGCGTCAAGAACGCGGCGAAAACCACTTCGACACCTCTTAT
GAGTTAGAATCACTCCTTGAGCAGCGTGAAGCGTCAACTGCTGGCGAAGTACAGTCCATCTGTCGCC
GGCGGTGACCATATTGAAACGTGCGTCAAGGGCGAACCTTAGGTTAGGCCACTCCATTGTTGCGCAG
CTGCCATTGGTGACAACCCATTGTCGTGGTACTGCCAGACGTTGATCGACGATGCCAGCGGCCACCCG
CTACGTTACAACCTTGCTGCCATGATTGACGTTCAACGAAACGGCGCAGCCAGGTGCTGGCAAAACG
TATGCCGGGTGACCTCTGAAATACTCCGTATCCAGACTAAAGAGCCGCTGGACCGTGAGGGTAAAGTCA
GCCGATTGTTGAATTATCGAAAACCGGATCAGCGCAGACGCTGGACTCAGACATCATGGCGTAGGT
CGCTATGTGTTCTGCCATATTGGCGGAACTGGAACGTAUTCAGCCTGGTGATGGGACGTATTCA
GCTGACTGATGCTATTGCCAGCTGGCGAAAAAACAAATCCGTTGATGCAATGCTGATGACCGGCACAGTT
ACGACTGCGGCAAAAAATGGGCTATATGCAAGCGTTGTAAGTATGGCCTACGCAACCTGAAAGAAGGG
GCGAAGTCCGTAAAGGTATTGAGAAGCTGTTAAGCGAATAATGAAAATCTGACCGGATGTAACGGTTGAT
AAGAAAATTATAACGGCAGTGAAAATTGCGAGCAAAAGTAATTGTTGCGAATCTTCTGCCGTTTTTA
TATAACCATTGAGAATAACAAACGAGTTAGCAGTAGGGTTTATTCAAAGTTTCCAGGATTTCTGTT
CCAGAGCGGATTGGTAAGACAATTAGCGTTGAATTTCGGGTTAGCGCAGTGGGTAACGCTCGTCAC
ATCATAGGCATGCAGTGCCTGGTAGCTGTAAGCCAGGGCGTAGCGTGCATTAATACCTCTATT
AATCAAACCTGAGAGCCGTTATTACAGCATGCTCTGAGTAATGGAATAATTAAAGCTAGCAAG
GATACTGTTACTGGTGGCGCAGGATTATTGGTTCTGCTGTTGTCGTCACATAATAAAATACGCAAG
ATAGTGTGTTAATGTCGATAAAATTACATACGCCGAAACCTGGAATCGCTCGTAAATTCTGATTCT
GAACGTTATTGAGCATGCAGATATCTGCGATGCCAAGCGATGGCTGTTACCGCACAGCACCA
GCCAGACGCGGTGATGCACCTGGCAGCAGAGAGCCACGTGACCGCTCAATAACTGCCCTGCGGCAATT
TTGAAAACCAATTGGGGTACTTATGTTCTTTAGAAGCGGCGCGCAATTATTGGTCTGGTCTGGATGAT
GAAAAGAAAAAAACTCCGCTTCATCATATTCTACTGATGAGGTGTTGGTACTACCCATCCGGA
TGAAGTAAATAGCAATGAAACGTTGCCGCTATTACGGAAATGACAGCATAACGCCAAGTAGTCCATATT
CTGCTTCTAAAGCTCCAGCGATCATTGGTTCGCGCATGGAAACGTAATTGTTACCGACCATTG
ACTAATTGCTCGAACAACTATGGCTTATCATTTCCGAAAAGCTTATTCCACTGGTATTCTAATGC
ACTGGAAGGTAAAGGCATTACCTATTATGGCAAAAGGAGATCAGATCCGCACTGGTGTATGTAGAGGATC
ATGCTCGAGCGTTATACCGTCGTAACCGAAGGTAAAGCGGGGAAACTTATAACATTGGTGGACACAAC
GAAAAGAAAACATCGACGTAAGTGTCACTATTGTTGATTTGGATGAGATAGTCCGAAAGAGAAATC
TTATCGTGAGCAAATTACCTATTGCTGATGCCCAAGGGCATGATGCCGTTATGCAATTGATGCCGATA
AAATTAGCCCGAATTGGGCTGGAAACCACAGGAAACGTTGAGAGCGGGATTGTAACACTGTGGAAATGG
TATCTGCCAATACAAATGGGTTGATAATGTAAAAGTGGTGCCTATCAATCGGATTGAACAGAACTA
TGGGGCCGCCACTAATGAATATCCTCCTTTGGAAAACAGGGCAGGTTGGGAAACTACAGCGTGC
TCTGGCACCTCTGGTAATTGATTGCTCTGATGTTCACTCCACTGATTACTGTGGTGTATTGTAACC
CTGAAGGTGTGGCTGAAACCGTTAGAAGCATTGGCCTGATATTATTGTAACCGCAGCGCTCACACCGCA
GTAGACAAAGCAGAACGAAACCGGAGTTGCACAATTACTGAACCGCAGCGTGAAGCGATCGCAGA
AGCAGCCAATGAAGTCGGCGCTGGGTATTCACTACTCTACTGACTACGTATTCCGGGGACCGGTGAAA
TACCATGGCAGGAGGAGGATGCAACCGCACCGCTAAATGTTACGGTAAACCAAGTTAGCAGGAAAGAAA
GCATTACAAGAGCATTGCGAAGCACCTTATTCCGGACCGAGCTGGTCTATGAGGTAAAGGAAATAA
CTTCGCCAAACGATGTTGCGTCTGGCAAAGAGCGTGAAGAATTAGCGTTATTGATCAGTTGGTG
CGCCAACCTGGCGCAGAGTTGCTGGCTGATTGACGGCACATGCCATTGCGTGGCACTGAATAACCGGAA
GTCGCAAGGTTGTACCATCTGGTAGCCAGTGGTACCAACCTGGCACGATTGCTGCGTGGTTTG
AGAGCGCGAAAGCAGGATTCCCTGCACTCAACAAAGCTCAACGCAAGTACAGCTATCCTA
CACCAAGCTCGTCGCCACATAACTCTGCCATTAAACAGAAAATTTCAGCAGAACTTGCCTGCTGTT
CCTGACTGGCAGGTTGGGTGAAACGCACTGCTCAACGAATTATTACGACTACAGCAATTAAAGT
GCATCTGTTGATGGGAAACAAGATGAATTAAAGGAATGATGGAATGAATACCGTAAAGGTATTA
TTTAGCGGGTGGTCTGGTACACGTTTATCTGTAATGGCTGTCAGTAAACAGCTGTTACCGATT
TATGATAAAACCGATGATCTATTACCGCTCTACACTGATGTTGGCGGGTATTGCGATATTGATTAT
CAGCACGCCACAGGATACTCCTCGTTCAACAACTGCTGGGTGATGGGAGCCAGTGGGGCTAAATCTC
ACTACAAAGTGAACCGAGTCCGGATGGTCTGCGCAGGCATTATCATGGTGAAGAGTTATCGGTGGT
GATGATTGCTTGGTACTGGTGATAATATCTACGGTCAGCACCTGCCATTGCTTAAGTTAATGGATGCC
TGTTAACAAAGAAAGTGGTCAACGGTATTGCTTACGTTAATGATCCTGAAACGCTATGGTGTGTTG
AGTTGATAAAAACGGTACTGCAATCAGCCTGGAAGAAAACGGTACAACCAAAAGTAATTGCGGT
ACCGGGCTTATTCTATGATAACTACGTTGGAAATGGCGAAAATCTTAAGCCTGCCCCGCGGTGA

ACTGGAAATTACCGATATTAACCGTATCTATATGGAACAGGGGCATTATCTGTTGCCATGATGGGACGTG
 GATATGCCTGGCTGGACACGGGGACACATCAAAGTCTTATTGAAGCAAGCAACTCATTGCCACCATTGAA
 GAGGCCAGGGCTGAAAGTTCTGCCCGAAGAAATTGCTTACCGTAAAGGGTTATTGATGCTGAGCA
 GGTGAAAGTATTAGCTAACCGCTGAAAAAAATGCTTATGGTCAGTATCTGCTAAAATGATTAAGGTT
 ATTAATAAAATGAATGTTATTAAACAGAAATTCCAGATGACTGATTTGAACCGAAAGTTGGTGAAT
 TGAGCGTGTTCTTATGGAAAGCTTAATCAGAAAGTTTCGAAGAGGCTGTAGGGCGGAAGGGTGAAT
 TTGTCAGGATAATCATTCTAAATCGTAAAGGTGACTTAGAGGTTACCTTCAGCTCCCTCCITT
 GAGCAGGCAAATTAGTAAGGTGATAGTTGGCGAGGTATTGATGTTGCACTAGACATTAGACCTAATT
 TGAAACATTGGTTCATGGGTTGGAGTAACTCTTCGTCAAGAAATAAAGGCAGCTATGGATTCCAGAAG
 GATTGCCCATGGTTTAACCTTAAGTGATATTGCAAGAGTTGTTATAAAACTAACAACTATTATTCT
 TTAATCATGAAAGGGAGTCATTGGAACGATGAGGAATTAAACATTGCCCTCTCAATCAGAGAA
 GATTCTGTACAGAAAGATATTAATTACCATCATTAGATTGTCAAATGTTAGCAAGTAGTGTATC
 TTTACACTGCACATAGTCATCATTCTATGCTTAAAGTAAATTATATTGCAACATCTATAACACAAAGCG
 AATAATATTGACCTGATGAAAGTTGTTATTATCTTCTAGGCCTTTATGACTAAAATAGTT
 GTGGTTCTACAGCTCAAATATTCCGACAAATAATGGGTACAAAGTTCTGTATTAGGAAGAATTGATGA
 GTTATTAAATGAGGATAATGAGGTGTTGATTGAAATAAACCTTGAAATGTTACGGAAAAGAAAGATG
 AATTAATACCAACAAGATTAATAATATTCAAAGATATGAAAGTAAAAAAATATCTAGATCATTATGCC
 GAGTACAAATATTATGATATCAGAACTCGTATGAAACAATTATTTCTCTGCTGACATTAGAGATAA
 CATAAAAAGATAATTGATTAGAAAAACCTCTATTATTGCTGAGTCTATATGGCGITGCAAGCAT
 TGCCTATTGAAATTAGTGCAGAAATACACTGTGTTATTGATGATGTGGCAACTGATTCTTAAAGAAATG
 TTTGATCTCATAATGAGGTTGACAAAAATTGTTTAAATGATTACCTAAAGTTGAAATTACTGA
 AGAAAATATTCAAACGTTGAGAGTTGAGCAATTATCTTCTGACAGAAGAAGATAATGTTGGTATA
 AAACAAGATAAAATATTGATGAGGGTTGTTGCTCTAGCGAGCAATCATTGTTAGAAAAGATTAAG
 AGAACTATCAATTCCAAACCCCTTCCTGCTTATTCCCGTAGCATGAAATTTCACAAAATTTCACGG
 CTTAAATTGGTTATAAAAATATATCCTGGATTAAATAGGAAAATAAGAATAGTTGIAACAGGAAAGG
 CATCAGATAAAAATAAAGATGTTAAACTGTGGAGAGGAAATTACCTTACGGGAGAGCTGACTTTCC
 ACATATAATAAAACTTAGCTCAACATGCTGTGTTATTGACCGATTACAACGGGACTGGAATTAAAAT
 AAAAATATTAGAAGCTGACAAAAGGTATTCTGACTTACAACAAAATTGCTTAAAGGAATATGTT
 CCGATTATGTTTATTGCGAGGAGGATACTGACACAAACTTGTCAATTAAACAGTTCTTGAA
 ACGACATTAAGAGTCCAAGAATGAATTATTGCTTTCTAGCTCTGCTTGGTTAATATTGGCTT
 GCCCATAATAATAAAAGTGGAGATATTAACGCATACTTAATGTTTCTGTTGCTTAATGGTATTAAAT
 ATCAGGGCTGCGTATGAATGATAGTGATTATATCGAATACAGGAAAATGTATAATGAAAGTGCCTATT
 GTGACTTCTGCGCATCTATAAGAGATATACATGGGAGGTAGGCTATCTATTCTATCATCAATCTT
 AAAACTTATGCTGCCATTCAATTATTCTTTTATTGCTTTTATCCTCTGCTTACATATT
 TTCATTGAGAAAATAAGTTAATACCGATACTATCGTAGTTTATTAAAGCATGCTTTATAGTT
 GAGATTGATTCAAATTAGGGCAGGATTAGCTGTTAGCATATCATTATATTCAATAATTAAAGGA
 AATAAAAGTATAATTACAGGAGTTTATTGCTTCTTGATTGCTTGGGGCGCTTATTATTGCTCTTG
 TTATCCTTTTCAAAAAAAACATAACATTAAAGTGTGTTTATTGCTTAAAGGCAATTATTAAAGG
 TTTCTATTGAAATGGGCTTAATTGATCGATACAACCTTATCTCAATATAGTTGCTTCAACTGCAATT
 TCGAATTATGTTGGTGGGAAGAATATGATTATCGGGTGAGTATATTACTAATCCGGTTTTATTAAAGG
 TGTTTTTAAATTGCTTAATGCACAAATATGACTTCACTGAGTATTTAAAGGAAATTATAGTGCTT
 ATAACCTTATATGTTAGGTGATTAGCTATGTTGCACTGAGTGGGATGGCTATTCTTCAGGCCGCTT
 TCATCCTTCTGACACTAGGTGAAAGCATTAAATTGATATGCTCTGTTCAACAAAGAAATACACCTCT
 GGCCTTCTAATTCTTAAACAATTGTCAATTAGGATATGATCTATTATTCTAATGTGCATC
 CTGAGCTTACTCTGATTATTTGGGTAATCTAAGTAAAAATAAAATAGGCATACTTATCTCTAA
 ATACAAAATCTGGACCTGTAATGTTAGTACGAGGATTGATAAAAGAAAATAAAATATGCTTTACTGT
 TTTTGTAAACAAATAGCGTAGATAAAATATATGATGAGTTATGCTGTTAGGAGCCAAGGTTATAT
 TAATACCGATGGTACTGGTTCAGCAAATTGAGAAGTTTAAAGGAACATCCACACATAAT
 ATCTTACATTACATGGGATCACGGCGATATGTTCTACTTCTGAATGGCGTAAAATATCTACTAT
 TCACAATAGACTAGATGAGGATTATATCCCATTATTGCGCGGTTAAAGGGAAATGCTATATATTCTC
 ATCGTTTATATTACGAAGATTAAATCATATCGTGTGCTCAGCAGCGGTCAATCAAACGAA
 TCGAAAGTAAAACCTAAACACCCATCCAGAAATGGGATTGATATAACTAGGTTAAGACACTGAGTC
 TGATAAAAAAAATTATTGAGGGAAAACACGGATTGATAGTGAAGAAAAGAAATATTATATTGTT
 CGTTATCATTAGGAAAATATTGCTTACCTCTGGAACACTTAGCCATGAAAGAAAATGATATATT
 ATTCTAGGTGATGGTGAACCTTTAGATATTGTAAGGATAAAATTCTAAAGGATTACGGTATATATT

GGGGAAAGTTGAATGCCCTTGAATATTATCAATTATCAGATATTTGTTCCGTTCTTATCGGAAG
GGCTCCCTTGCACATTAGAAGCTGCCTACTGGGTGCTATTATGTTAGCGATATAGAGCCCCAT
AGAGAAATTGCATCTATTAGGAGAGGAAAATATTCTATGTTAAAATTAAGGATGGATCATATAATT
TTGCAACCTAAAATAAAAAGCTGACTATAACGCTTTCTGACGATAAACCTACAATATATCCGATA
AAAAAATGTCAAATCTTATGACAAACTTTGTTCTTATTAGAGCAGAGGACTAATATAATGATT
TGTTCGGTATTCTCATGGTCATTCAAAACTCTAAGGAATTAGGAGCAGTATCAAATTAAATAATC
ACAGCAGAATTAAAGTTATCATCAAAGATAATTAGGAGAGCAGCAGTATTGGATTGGTCAGGAAAC
AAAATAACTTATTAAAGGCTAAAGAGAAAAGGATTGGAGAGAATAATAATGAAGTTTCTCTAT
ATCCTCCTTAATTACTAAGGAAGATTGGTTATGAATCCTGATATATATTGAGTGCTCTGATC
TATTAGATGTCGTAGATGAGTGTGGTCAGCGAATGTTAATCTAGCAACGATAAATTATACAGGGATT
GATAAAAACATATGATAACTCAGTAAGGAAATTCCCTCGGCAATTGATTTTATGTCATTATT
TAAGAAAATGACTGTAGTAAATAAGAACAAAATAACGAAACACATATGTTGATTGGCTGCAGGTT
CTTTCTAATATTAAATGCCCTCTTTATTCAAAACTCAACGGATTCAACGAAAGTATTATGTTGC
GAAGATATTGATATGTTGGCGAGCTAAAACACTTCAACTTCAGTTTAACTATCCATGCTATGC
AGCAATTCAATTGGCACAAATTAAACATCGTAGGATTTAGTAGACATTCAATTGGCATATAAAAGTA
TTATCCTTTTATTATATAAAATGGTATGCTGCGTTCTAGTAAGTTGCTTAATGCTAATATTCTTT
AAGAGGTGAGAATGATACTGTTATTGGCTGGTGGTCGGGAAGTCGTTGAGCCACTTCACGAGAA
AAGTCCCCAAGCAGTTTAAAGTTGACTGGCAGTTGACAATGTTGCACTGTCAGTCAACATTGTCACGTCTAA
TAATTAAATGCTGATGATTCAATAGTTATGCAACGAAGAGCATAGATTATTGTTGCAAGAACATTAA
GAGAGTTAGGCAAACCTTCAAATAACATTATTCTGAACCCAAAGGTCGTAATACAGCCCCTGCTATAACA
CTCGCAGCATTAGCAGCAAAAGAAAATCGCTGATGAAGATCCATTGATTCTTATTTAGCTCAGATCA
CAACATCCAAGACGAACATGTTCTGTGAGGCAATTAAAGGCGTCATCTTAGCTAGTTATGGAAAAC
TAGTGACTTTGGTATGTTCCATTCAAACCTGAAACTGGGTATGGCTATATTGTCGCGGTGATGAAGTG
CCTGTAGATGAGCAGCATGCGGTGGCTTGAAGTGGCGCAGTTGTCGAAAACCGAATCTGAAACCGC
GCAGGCCTATGTCAGCGCGAATTACTGGAACAGCGGTATGTTCTGTCGCGGTGACGCTATC
TCGAAGAACTGAAAAGTATGTCGGATATTCTGATGCCGTGAAAAAGCGATGAGGCCGTCGATCCG
GATCTGATTATTGTTGAGGCGTTCTGCTGGCAAGAGTCGGTGGATTACGCGGT
CATGGAATGCAAGGCAGATGCCGTTGTCGGCGATGGATGCGGGCTGGAGCGATGTCGGTCTGGCTT
CATTATGGGAGATCAGGCCAACACCGCCGAGGGCAACGTTGCCCCAGGGCGATGTGATTATCACAAA
GAAACAGCTATGTAACGCCGAACTGGCTGGTCAACGCCGTTGAGGCGATGTGAAAGGTTGGTAGTGCA
GACCAAAGATGCACTGTCGATTGCCGACCGTAATGCCGTCAGGATGTGAAAGAAAGTGGTCAGCAGATCA
AAGCTGATGGTCGCCATGAGCATGGGTGCATCGCAAGTGTATGTCGTGGGCAAATATGACTCTAC
GACGCCGGCAGCGTACAGGTGAAACGCATACCGTGAACCGGGCAAGGTTGTCGGTACAGATGCA
TTATCATCGCGGAAACACTGGGTGTTGTCGCGGGAAACGGCAAAAGTCACTATCAACGGTATATCAAAC
TGCTGGTAAACAGAGTCCATTATATTCCGCTGGGGCGATGCACTGCCTGGAAAACCCGGGAAATA
GATTAGAATTAAATTGAAGTTGCTCTGGTCATATCTGAAGAAGATGATGTTATTAGATGTTATGATCG
CTATGGACGAAAGTAATATATAATTATTCAGAATTAGAAATGATAATTATAAGTTGCTGGATA
AACAAATAGATAGTATGGGTGGAAAATATGAGTTCTTAACTGTTAAAGCTTACGACATTGCGGGAA
ATTAGGTGAAGAACTGAATGAAGATATGCCCTGGCGCATGGATGTCGCGCTATGGCAATTCTCAAACCGA
AAACCATTTGTTAGGCGGTGATGCCGTCTACCGCGAACCTTAAACTGGCGCTGGCAAAGGTTA
CAGGATGCCGGCGTCACTGGATATTGGCATGCCGACCGAAGAGATTATTCGCCACGTTCCA
TCTCGCGTGGATGCCGCAATTGAAGTTACGCCAGCCATAATCCGATGGATTACAACGGCATGAAGCTGG
TGCAGCGAAGGGGCTGCCCGATCAGCGGTGATACCGACTGCCGACGTCCAGCGTCTGGCAGAAGCTAAC
GACTTCCTCCCGTCAAGAACCAAACCGCGTCGCTACGCAAATCAATCTGCGTACGCTTACGTTGA
TCACCTGTTGCTTATATCAATGTCAAAACCTTACGCCGTCAGCTGGTATCAACTCCGGAAATGGCG
CAGCGGGTCCGGTGGACGCTATCGAAGCCGCTTAAAGCCCTGGCGACCGTGGAGTTAATCAA
GTGCATAACACGCCGGACGGCAATTCCCAACGGTATTCTAACCGTGTGCCGAATGTCGCGACGA
CACCCGCAATGCCGTACAAACACGGCGGGATATGGGCATTGCCATTGATGGCGATTGACCGCTGTT
TCCTGTTGACGAAAAGGGCAGTTATTGAGGGCTACTACATTGTCGGCTGCTGGCAGAAGCGTCTC
GAAAAAAATCCCGCGCGAAGATCATCCACGATCCACGCTCTCCCTGGAACACCATTGATGTGGTGACGGC
CGCAGGGCGACGCCGGTATGTCGAAACAGGACACCGCTTATTAAAGAACGTATGCGCAAGGAAGACG
CCATCTACGGTGGCGAAATGAGCGCTCACCATTAACGCCGATTTGCTTACTGTCAGCGCATGATC
CCGTGGCTGCTGGTCGCCAAGTGGTGTGGCTGAAAGGAAAACGCTGGCGAAGTGGTGCACCGGAT
GGCGCGTTCCGGCAAGCGGTGAGATCAACAGAAAATGCCGACCCGTTGAGGCAGTTACCGCGTGG
AACAGCATTAGCCGTGAGGTGCTGGCGGTGGATCGCACCGATGGCATGACCTTGCCGACTGG

CGCTTAACCTGCGCTTCCAACACCGAACCGGTGGTGCCTGAATGTGGAATCTCGCGGTATGTTCA
GGTTATGGAATCCATACTCAAGAAATATTATCAATTGACGTCAAAGAATAAGCCTGACAAGTTAG
GGCTTAATTAAATATATTTTGAAATTGGGATTTGGTAAGATTAAATATGTTATTAATGTTAATGTT
TTGAATTAAATGTTGACTGGAAAATAATAATGAGAACGAAAAAGCATTACACAACCTTAAAGTTGATT
TAATTACTTTTATTGGTTGCTAGGGTTATATTGAACTGTTGTTCAAAATGGGAAAGTGT
ATTACTGGAGTGTACTATTCACACAGTGACAGCATCTCAATTGGCAGAATTAGGTATTGGAAT
TGCAGCTGCCAGCGTATTATATAACCGCTCAGCGAGAATGAATAACAAATAACTACATAATCTT
TGCTCTCAGTCATATAACAAATATAATTGTGTTGTTGATTCTGGCGTGTATAGGTATCTGTATT
TATTACTTTATTGATTCTGAAAGGGTGTAAATGGCGTTTTTATATTGGGCTTGTGTTAATAC
ATCGTTGACATATAGTTATGCTAAATACCTCACATTAACTGCTAATCAGCGGTACTCAGCAGTAAGAA
AAATTCAAGGTGGCGAAAAGTTATAATAATTGTTAGTATTCAAGATATTAAATTGCTTACGCAAAGTTTC
ATACTTATTGTTAGTTGAGACTTAGGTATTCTCAATATTGATTTAAAAAAATAATTGGGAA
CGGAAATCAATATCTCAGTAATGAGGTTACTTATTGAAAGCGATAAAACTTTGATAAAAAAGAATTAA
AAATAAGAATAAAAATGTTCTCCATAAAATAGGTGCTGTGCTGCCTTAATACAGACTACCTGCTT
GTATCAAAGTTCTGACATTAAGTTATGTGACAATTGGCAGCTATATGATGGTATTCAGATAGTAAC
TGTTTGATGTCAGTTGTTAATGCTATTACTGCAGGAATGGTAATTACTTAATTAATAAAAGTAATT
TAGAAATTAAAGGAAATTACACGTCAATTATGATATTATCGCCTTGCAACATTCACTAAAT
ATGTTTTCTGTTAATGATTATCGAAATGGTAGGTGTTAATTACATTAAGTAACACCCCTAGT
TGCATTAATGATTGTTAACGTATTCAATTAGTGTGTCAGGGTACCTCTGATATATTAAAAACGCAAGTG
GACATTGGTGTATTTATTCCATTATTAGAAGGTGTGCTGAATATTACGATATCCATATTGGCT
ATCATTATTGGATTACCTGGCATTATTAGGGACAATAGTATCTAATTAGTAATAATGCTTGC
ACCATTATATCTTACTCTAAGTTATTAAATCTTAGAAATCCGACGAGGGTTATTGAAATTATTCTC
GGCCTATGTTATTCATTATGTGATTGGGGTGAGCTATTATTGCGATGAAATATTCAATTAA
GTAAGTACATGGTGGATTAAACAGCTACTCTTAGTCTACTCCTAGCATATTGTAATATGTG
TATTCTCTACGGATAGTGACTTAGATTATTTCAGAAAATTATATATGATTATGAAGAAATAAA
AATTGAAAATGTTAAATCGAAATTATGCAACGAGCTTATTGAAATTGATATGTGATTTTC
GAATAGGAGTAAGGATCCGTGAGGCTGGAGCTCGTCAGGTTCTACTTCTAGAGAATAGGAACCT
CGGAATAGGAACTAAGGAGGATATTCAATGGATAAAGCCGTAAGCATAAGCATGGATAAGCTATT
ACTTTAATAAGTACTTTGTTACTTATTGCAACATTCCAGGCCGAGCATTAGCGCGGTGATCAC
CTGACAGGAGTATGTAATGTCAGCAACAGATCGCGTAGTCGGTATGGCAGTGATGGACGCAACCT
CGCTCAACATCGAAAGCCGGTTATACCGTCTCTATTCAACCCTGGTGGAGAAGACGGAAGAAGTG
ATTGCCGAAAATCCAGGCAAGAACTGGTCCCTACTACGGTGAAGAGTTGTCGAATCTGGAAAC
GCCTCGTCGCATCCTGTTAATGGTGAAGCAGGTGCAGGCACGGATGCTATTGATTCCCTCAAACCAT
ATCTCGATAAAGGAGACATCATATTGATGGGGTAACACCTTCTCAGGACACTATTGTCGTAATCGT
GAGCTTCAGCAGAGGGCTTAACTCATCGGTACCGGTGTTCTGGCGGTGAAGAGGGGGCGCTGAAAGG
TCCTCTATTATGCGTGGGCCAGAAAAGCCTATGAATTGGTAGCACCAGTACCTGACCAAAATCGCG
CCGTAGCTGAAGACGGTGAACCATCGTTACCTATATTGGTCCGATGGCGCAGGTCACTATGTGAAGATG
GTTCACACGGTATTGAAATACGGCGATATGCACTGATTGCTGAAGCTTACTCTGCTTAAAGGTGGCT
GAACCTCACCAACGAAGAACTGGCGCAGACCTTACCGAGTGGAAATACGGTGAACTGAGCAGTTACCT
TCGACATCACCAAAAGATATCTTACCAAAAAAGATGAAGACGGTAACTACCTGGTGTGATCCTGGAT
GAAGCGGCTAACAAAGGTACCGGTAATGGACCGAGCCAGCGCCTGGATCTGGCGAACCGCTGCG
GATTACCGAGTCTGTGTTGACGTATATCTCTCTGAAAGATCAGCGTGTGCGCATCTAAAGTTC
TCTCTGGTCCGCAAGCACAGCCAGCAGCGACAAGGCTGAGTTCATGAAAAGTTGTCGTGCGTGT
CTGGCAAATCGTTCTACGCCAGGGCTCTCAGCTGCGTGTGCGTGAAGAGTACAACGG
TCTGAACACGGCAGAACATCGCAAGATTCCCGTGTGCGTGCATCATCCGTGCGCAGTTCTGCAGAAA
TCACCGATGTTATGCCAAAATCCACAGATCGCTAACCTGTTGCTGGCTCCGTACTTCAAGCAAATTGCC
GATGACTACCAGCAGGCCGCTGCGTGTGCGTGAAGAGTACAAGCGTATTCCGGTCCGACCT
CTCCGAGCGGGTGCCTATTACGACAGCTACCGTGTGCTGTTCTGCCCTGCGAACCTGATCCAGGCACAGC
GTGACTATTGGTGCCTATTACGACAGCTACCGTGTGCTGTTCTGCCCTGCGAACCTGATCCAGGCACAGC
TAA

CLAIMS

1. A method of preparing a bioconjugate of an *E. coli* O_x antigen polysaccharide covalently linked to a carrier protein, the method comprising:

(i) providing a recombinant host cell comprising:

- a. a nucleotide sequence of an *rfb* gene cluster for the O_x-antigen polysaccharide;
- b. a nucleotide sequence encoding the carrier protein comprising at least one glycosylation site comprising a glycosylation consensus sequence having SEQ ID NO: 1, preferably having SEQ ID NO: 2; and
- c. a nucleotide sequence encoding an oligosaccharyl transferase PglB_y; and

(ii) culturing the recombinant host cell under conditions for production of the bioconjugate,

wherein:

when the O_x- antigen is O1A antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V;

when the O_x-antigen is glucosylated O4 antigen polysaccharide, the PglB_y comprises the amino acid mutation N311V or the amino acid mutations Y77H and N311V, and the recombinant host cell further comprises a sequence encoding a glucosyltransferase GtrS having at least 80% identity to SEQ ID NO: 4 and being capable of modifying an *E. coli* O4 antigen polysaccharide by addition of glucose to produce the *E. coli* glucosylated O4 antigen polysaccharide, and nucleotide sequences encoding a translocase GtrA and a glucosyltransferase GtrB having at least 80% sequence identity to SEQ ID NOs: 7 and 8 respectively, wherein the translocase is capable of translocating bactoprenol linked glucose and the glucosyltransferase is capable of glucosylating bactoprenol;

when the O_x-antigen is O6A antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V;

when the O_x-antigen is O8 antigen polysaccharide, the PglB_y comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483 and 669;

when the O_x-antigen is O15 antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V;

when the O_x-antigen is O16 antigen polysaccharide, the PglB_y comprises the amino acid mutations of Y77H, S80R, Q287P, K289R, and N311V;

when the O_x-antigen is O18A antigen polysaccharide, the PglB_y comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483 and 669; and

when the O_x-antigen is O75 antigen polysaccharide, the PglB_y comprises the amino acid mutation of N311V,

wherein in each case the amino acid mutations are relative to the wild-type PglB having the amino acid sequence of SEQ ID NO: 6, and

wherein the O1A, glucosylated O4, O6A, O8, O15, O16, O18A, and O75 antigen polysaccharides have the structures of Formulas (O1A), (O4-Glc+), (O6A), (O8), (O15), (O16), (O18A), and (O75), respectively, as shown in Table 1, and each n is independently an integer of 1 to 100, preferably 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20.

2. The method of claim 1, wherein the O_x-antigen is glucosylated O4 antigen polysaccharide, and the PglB_y comprises the amino acid mutation N311V or the amino acid mutations Y77H and N311V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

3. The method of claim 2, wherein the recombinant host cell further comprises a sequence encoding a GtrS having the amino acid sequence of SEQ ID NO: 4, and nucleotide sequences encoding a GtrA and a GtrB having the amino acid sequences of SEQ ID NOs: 7 and 8, respectively.

4. A method of preparing a bioconjugate of an *E. coli* O_x antigen polysaccharide covalently linked to a carrier protein, the method comprising:

(i) providing a recombinant host cell comprising:

- a. a nucleotide sequence of an *rfb* gene cluster for the O_x-antigen polysaccharide;
- b. a nucleotide sequence encoding the carrier protein comprising at least one glycosylation site comprising a glycosylation consensus sequence having SEQ ID NO: 1, preferably having SEQ ID NO: 2; and
- c. a nucleotide sequence encoding an oligosaccharyl transferase PglB_y; and

(ii) culturing the recombinant host cell under conditions for production of the bioconjugate,

wherein the PglB_y comprises the amino acid mutation N311V relative to the wild-type PglB having the amino acid sequence of SEQ ID NO: 6,

wherein the O_x- antigen is O1A antigen polysaccharide, glucosylated O4 antigen polysaccharide, O6A antigen polysaccharide, O15 antigen polysaccharide, O16 antigen polysaccharide, or O75 antigen polysaccharide,

and when the O_x-antigen is glucosylated O4 antigen polysaccharide, the recombinant host cell further comprises a sequence encoding a glucosyltransferase GtrS having at least 80% identity to SEQ ID NO: 4 and being capable of modifying an *E. coli* O4 antigen polysaccharide by addition of glucose to produce the *E. coli* glucosylated O4 antigen polysaccharide, and nucleotide sequences encoding a translocase GtrA and a glycosyltransferase GtrB having at least 80% sequence identity to SEQ ID NOs: 7 and 8, respectively, wherein the translocase is capable of translocating bactoprenol linked glucose and the glycosyltransferase is capable of glucosylating bactoprenol, and

wherein the O1A, glucosylated O4, O6A, O15, O16, and O75 antigen polysaccharides have the structures of Formulas (O1A), (O4-Glc+), (O6A), (O15), (O16),

and (O75), respectively, as shown in Table 1, and each n is independently an integer of 1 to 100, preferably 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20.

5. The method of any one of claims 1 to 4, further comprising isolating the bioconjugate from the recombinant host cell.

6. The method of any one of claims 1 to 5, wherein the carrier protein is selected from the group consisting of detoxified Exotoxin A of *P. aeruginosa* (EPA), *E. coli* flagellin (FliC), CRM197, maltose binding protein (MBP), Diphtheria toxoid, Tetanus toxoid, detoxified hemolysin A of *S. aureus*, clumping factor A, clumping factor B, *E. coli* heat labile enterotoxin, detoxified variants of *E. coli* heat labile enterotoxin, Cholera toxin B subunit (CTB), cholera toxin, detoxified variants of cholera toxin, *E. coli* Sat protein, the passenger domain of *E. coli* Sat protein, *Streptococcus pneumoniae* Pneumolysin, Keyhole limpet hemocyanin (KLH), *P. aeruginosa* PcrV, outer membrane protein of *Neisseria meningitidis* (OMPC), and protein D from non-typeable *Haemophilus influenzae*,

preferably wherein the carrier protein is detoxified exotoxin A of *Pseudomonas aeruginosa* (EPA), preferably wherein the EPA carrier protein comprises 1-10, preferably 2-4, more preferably 4, of the glycosylation sites, preferably wherein each glycosylation site comprises a glycosylation consensus sequence having SEQ ID NO: 2, preferably wherein the EPA carrier protein comprises SEQ ID NO: 3.

7. The method of any one of claims 1-6, wherein the recombinant host cell is an *E. coli* cell, e.g. an *E. coli* K-12 strain, such as strain W3110.

8. A bioconjugate produced by the method of any one of claims 1-7.

9. The bioconjugate of claim 8, wherein the bioconjugate is a bioconjugate of *E. coli* glucosylated O4 antigen polysaccharide covalently linked to a carrier protein, preferably wherein the carrier protein is an EPA carrier protein comprising SEQ ID NO: 3, preferably wherein the glucosylated O4 antigen polysaccharide has the structures of Formula (O4-Glc+) as shown in Table 1, and n is an integer of 5 to 40.

10. A composition comprising a bioconjugate of claim 8 or 9, preferably further comprising one or more conjugates each comprising an *E. coli* antigen polysaccharide covalently coupled to a carrier protein.

11. A recombinant host cell for preparing a bioconjugate of an *E. coli* O_x antigen polysaccharide covalently linked to a carrier protein, the recombinant host cell comprising:

- a. a nucleotide sequence of an *rfb* gene cluster for the O_x-antigen polysaccharide;
- b. a nucleotide sequence encoding the carrier protein comprising at least one glycosylation site comprising a glycosylation consensus sequence having SEQ ID NO: 1, preferably having SEQ ID NO: 2; and
- c. a nucleotide sequence encoding an oligosaccharyl transferase PglB_y

wherein:

when the O_x- antigen is O1A antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V;

when the O_x-antigen is glucosylated O4 antigen polysaccharide, the PglB_y comprises the amino acid mutation N311V or the amino acid mutations Y77H and N311V, and the recombinant host cell further comprises a sequence encoding a glucosyltransferase GtrS having at least 80% identity to SEQ ID NO: 4 and being capable of modifying an *E. coli* O4 antigen polysaccharide by addition of glucose to produce the *E. coli* glucosylated O4 antigen polysaccharide, and nucleotide sequences encoding a translocase GtrA and a glucosyltransferase GtrB having at least 80% sequence identity to SEQ ID NOs: 7 and 8 respectively, wherein the translocase is capable of translocating bactoprenol linked glucose and the glucosyltransferase is capable of glucosylating bactoprenol;

when the O_x-antigen is O6A antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V;

when the O_x-antigen is O8 antigen polysaccharide, the PglB_y comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483 and 669;

when the O_x-antigen is O15 antigen polysaccharide, the PglB_y comprises the amino acid mutations of N311V, K482R, D483H, and A669V;

when the O_x-antigen is O16 antigen polysaccharide, the PglB_y comprises the amino acid mutations of Y77H, S80R, Q287P, K289R, and N311V;

when the O_x-antigen is O18A antigen polysaccharide, the PglB_y comprises no amino acid mutations at positions 77, 80, 287, 289, 311, 482, 483 and 669; and

when the O_x-antigen is O75 antigen polysaccharide, the PglB_y comprises the amino acid mutation of N311V,

wherein in each case the amino acid mutations are relative to the wild-type PglB having the amino acid sequence of SEQ ID NO: 6, and

wherein the O1A, glucosylated O4, O6A, O8, O15, O16, O18A, and O75 antigen polysaccharides have the structures of Formulas (O1A), (O4-Glc+), (O6A), (O8), (O15), (O16), (O18A), and (O75), respectively, as shown in Table 1, and each n is independently an integer of 1 to 100, preferably 3 to 50, e.g. 5 to 40, e.g. 7 to 25, e.g. 10 to 20.

12. The recombinant host cell of claim 11, wherein the O_x-antigen is glucosylated O4 antigen polysaccharide, and the PglB_y comprises the amino acid mutation N311V or the amino acid mutations Y77H and N311V relative to wild-type PglB having the amino acid sequence of SEQ ID NO: 6.

13. The recombinant host cell of claim 12, wherein the recombinant host cell further comprises a sequence encoding a GtrS having the amino acid sequence of SEQ ID NO: 4, and nucleotide sequences encoding a GtrA and a GtrB having the amino acid sequences of SEQ ID NOs: 7 and 8, respectively.

14. The recombinant host cell of any one of claims 11 to 13, wherein the carrier protein is selected from the group consisting of detoxified Exotoxin A of *P. aeruginosa* (EPA), *E. coli* flagellin (FliC), CRM197, maltose binding protein (MBP), Diphtheria toxoid, Tetanus toxoid, detoxified hemolysin A of *S. aureus*, clumping factor A, clumping factor B, *E. coli* heat labile enterotoxin, detoxified variants of *E. coli* heat labile enterotoxin, Cholera toxin B subunit (CTB), cholera toxin, detoxified variants of cholera toxin, *E. coli* Sat protein, the passenger domain of *E. coli* Sat protein, *Streptococcus pneumoniae* Pneumolysin, Keyhole limpet hemocyanin (KLH), *P. aeruginosa* PcrV, outer membrane protein of *Neisseria meningitidis* (OMPC), and protein D from non-typeable *Haemophilus influenzae*, preferably wherein the carrier protein is detoxified exotoxin A of *Pseudomonas aeruginosa* (EPA), preferably wherein the EPA carrier protein comprises 1-10, preferably 2-4, more preferably 4, of the glycosylation sites, preferably wherein each glycosylation site comprises a glycosylation consensus sequence having SEQ ID NO: 2, preferably wherein the EPA carrier protein comprises SEQ ID NO: 3.

15. The recombinant host cell of any one of claims 11 to 14, wherein the recombinant host cell is an *E. coli* cell, e.g. an *E. coli* K-12 strain, such as strain W3110.

1/12

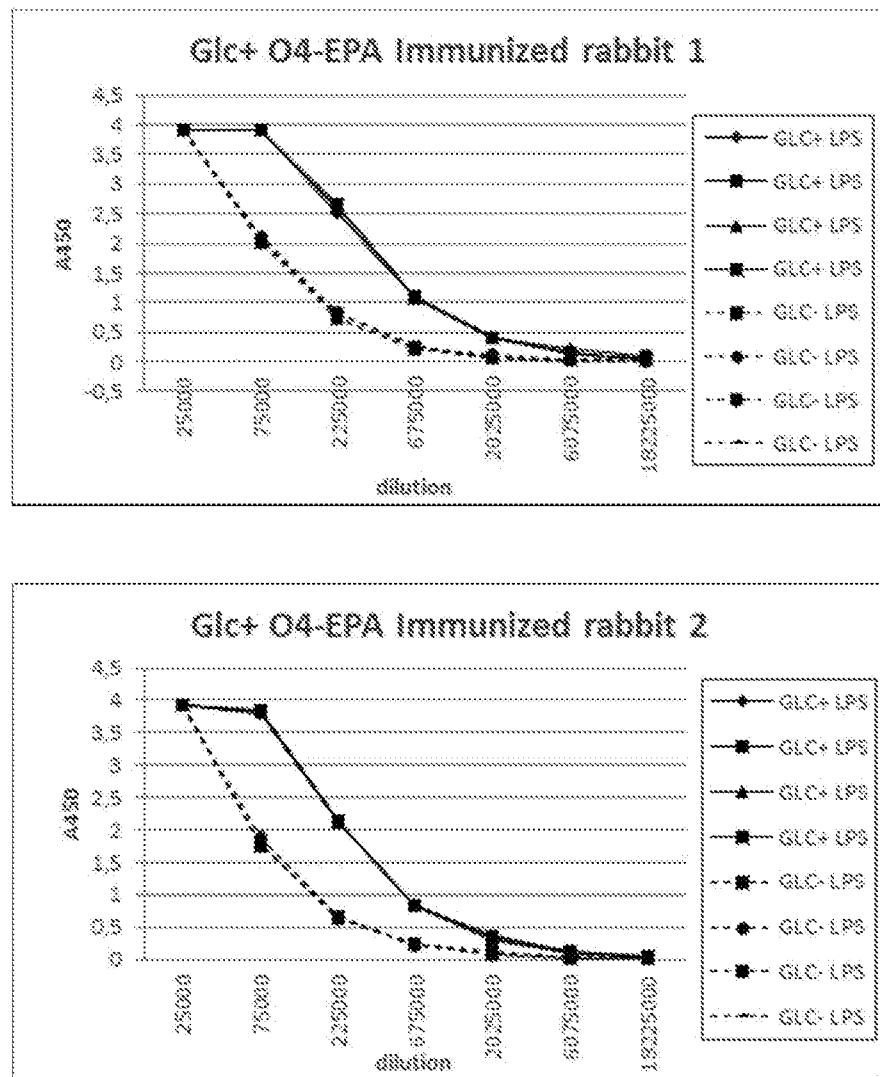


Fig. 1

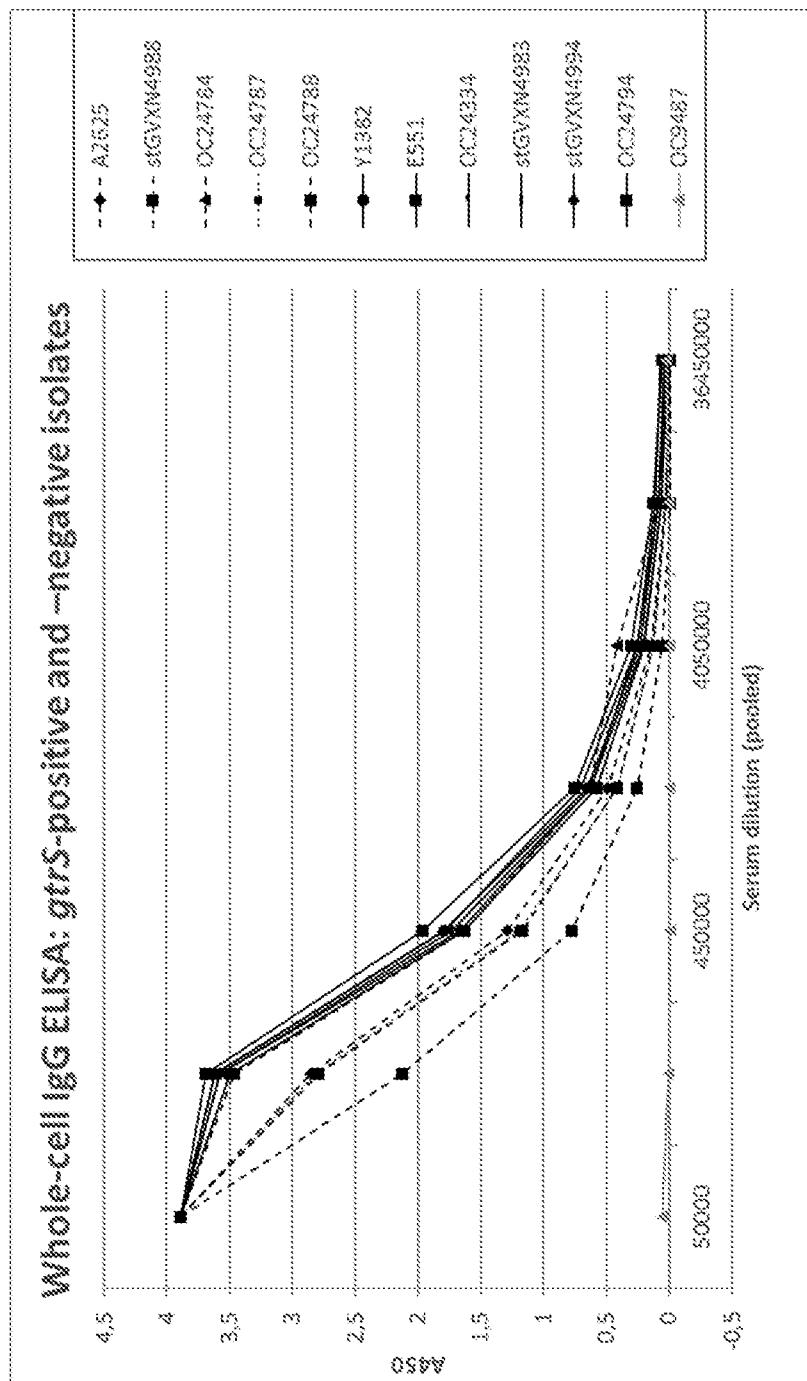
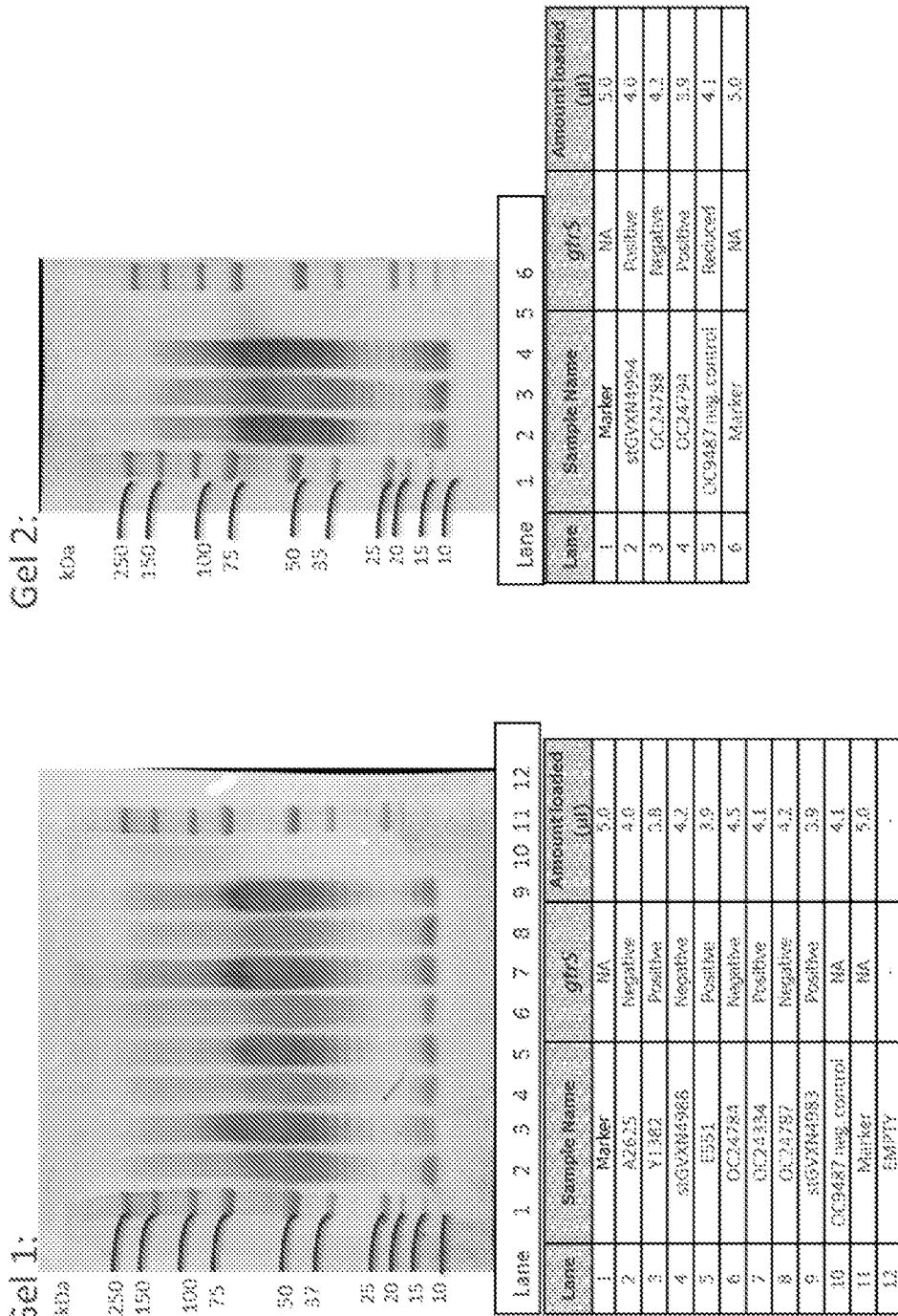


Fig. 2

3/12

Fig. 3



4/12

Fig. 4A

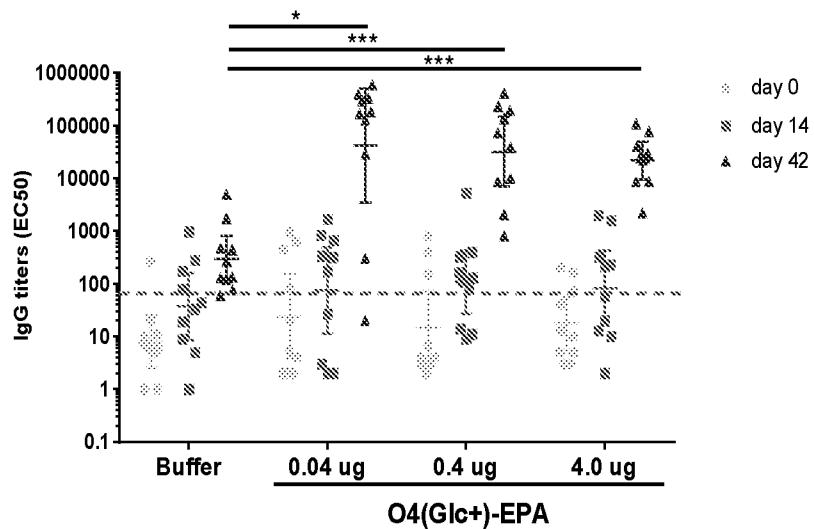
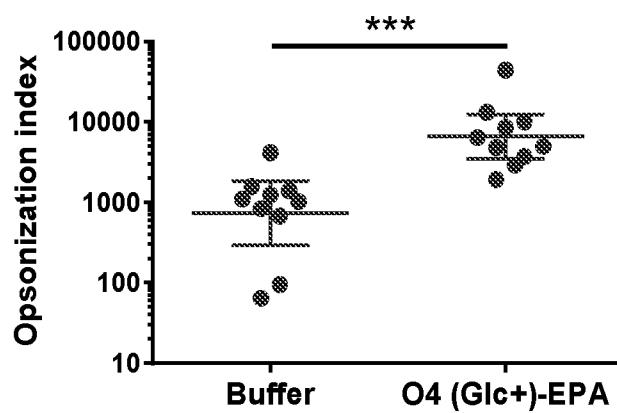


Fig. 4B



5/12

Fig. 5

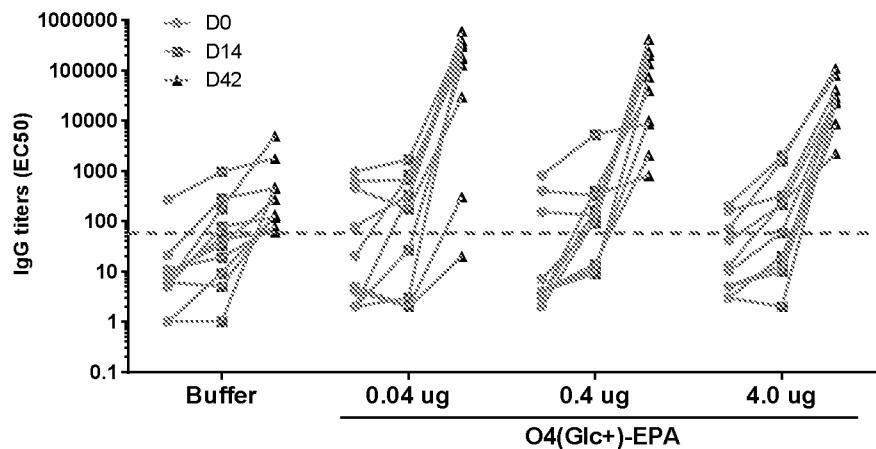
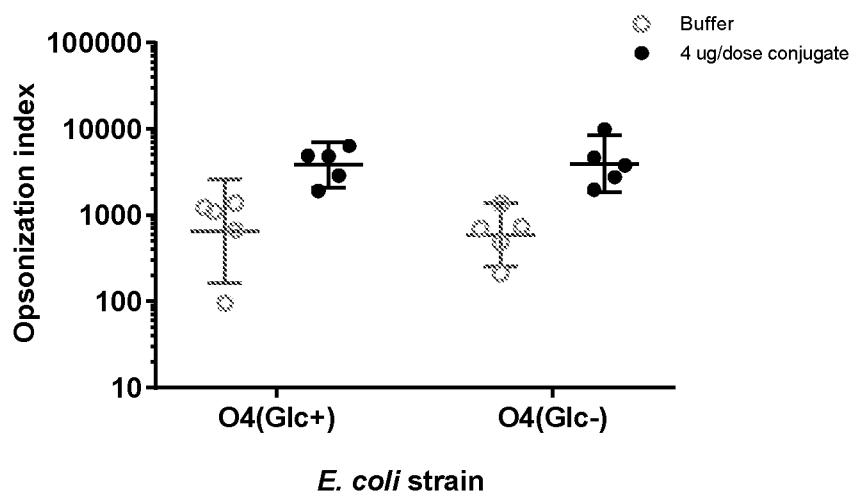


Fig. 6



6/12

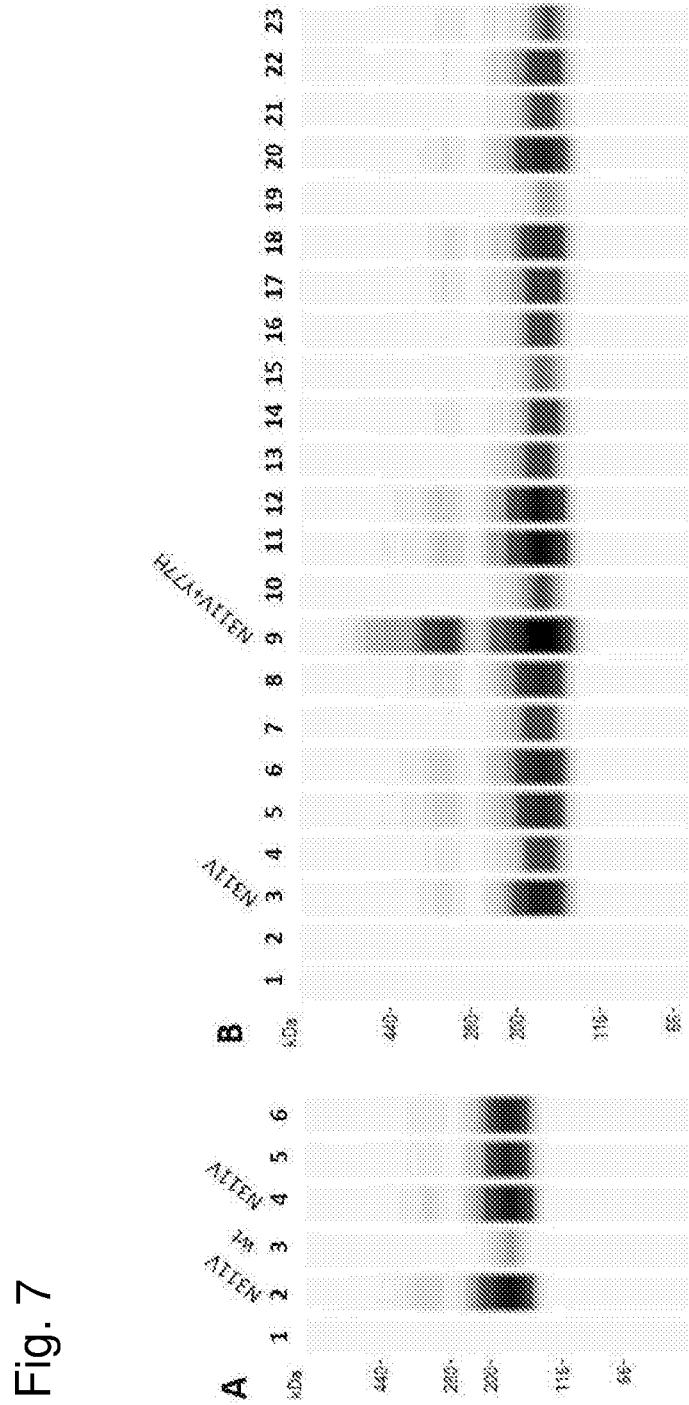


Fig. 7

7/12

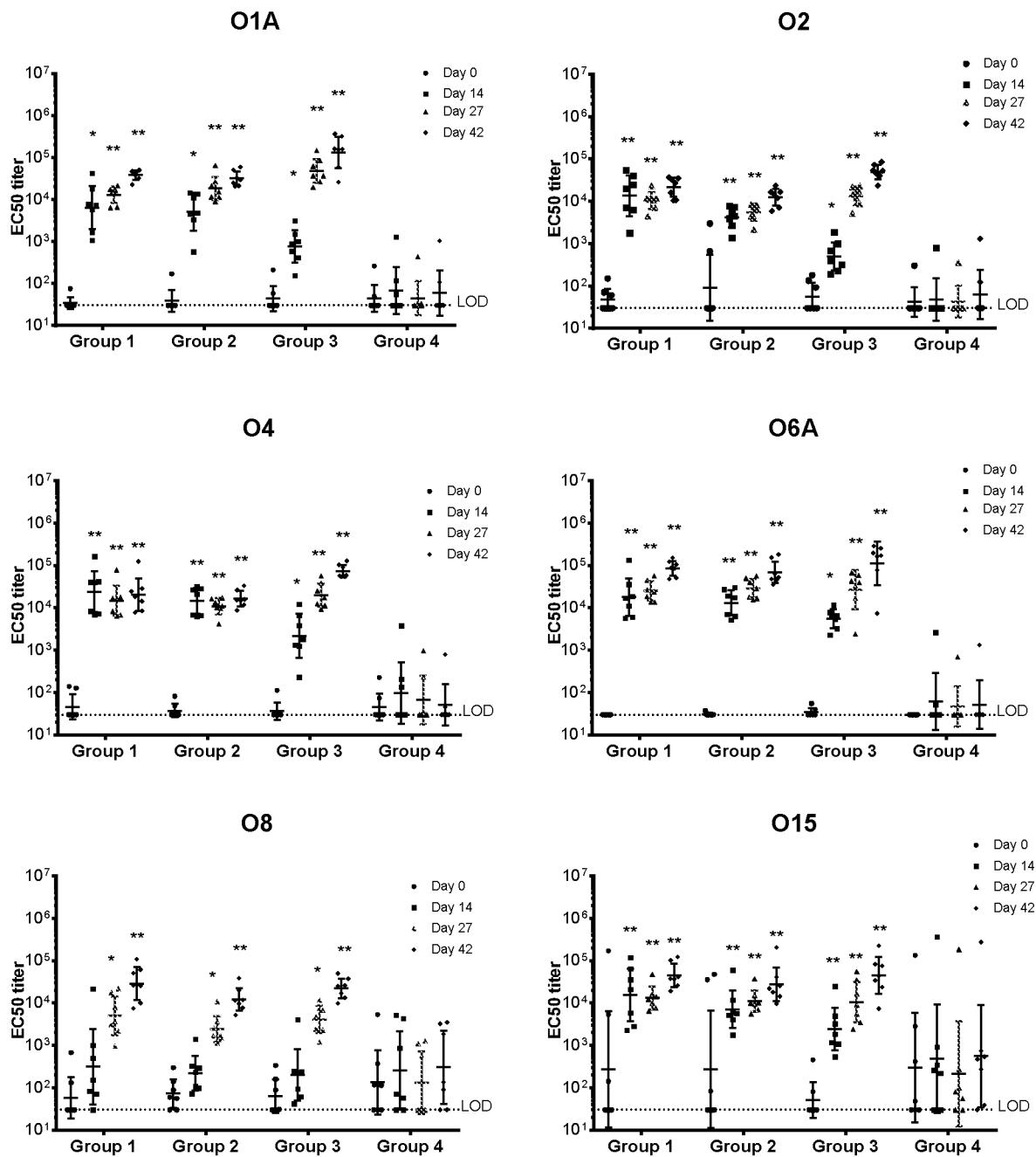
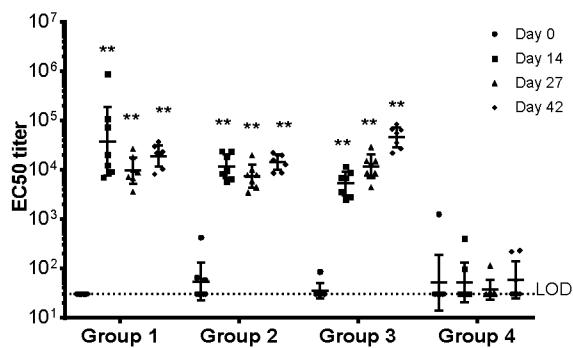


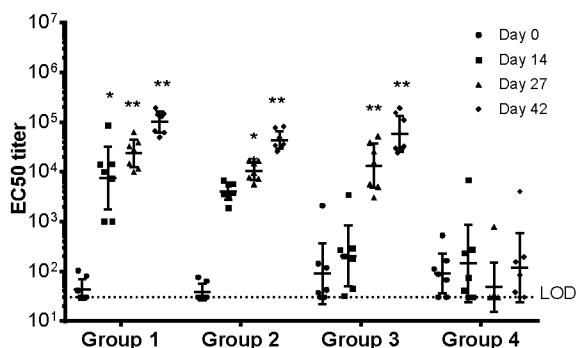
Fig. 8

8/12

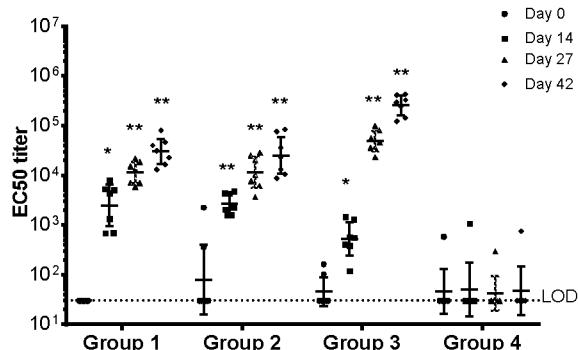
O16



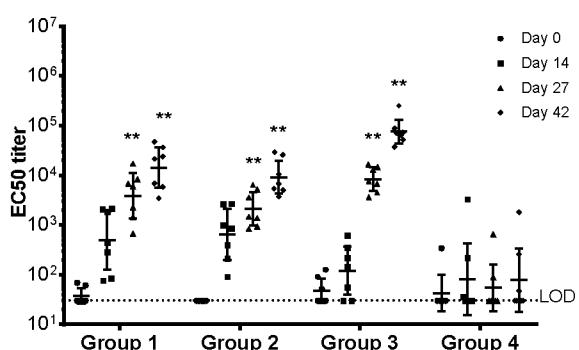
O18A



O25B



O75



EPA

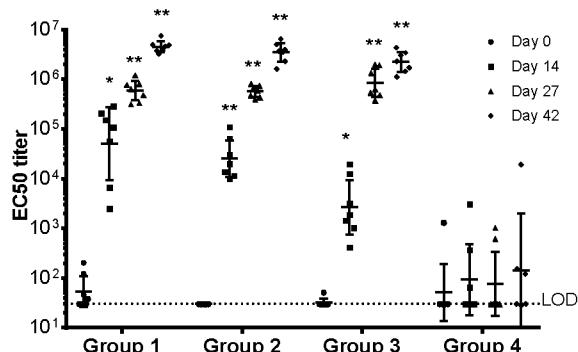


Fig. 8 - continued

9/12

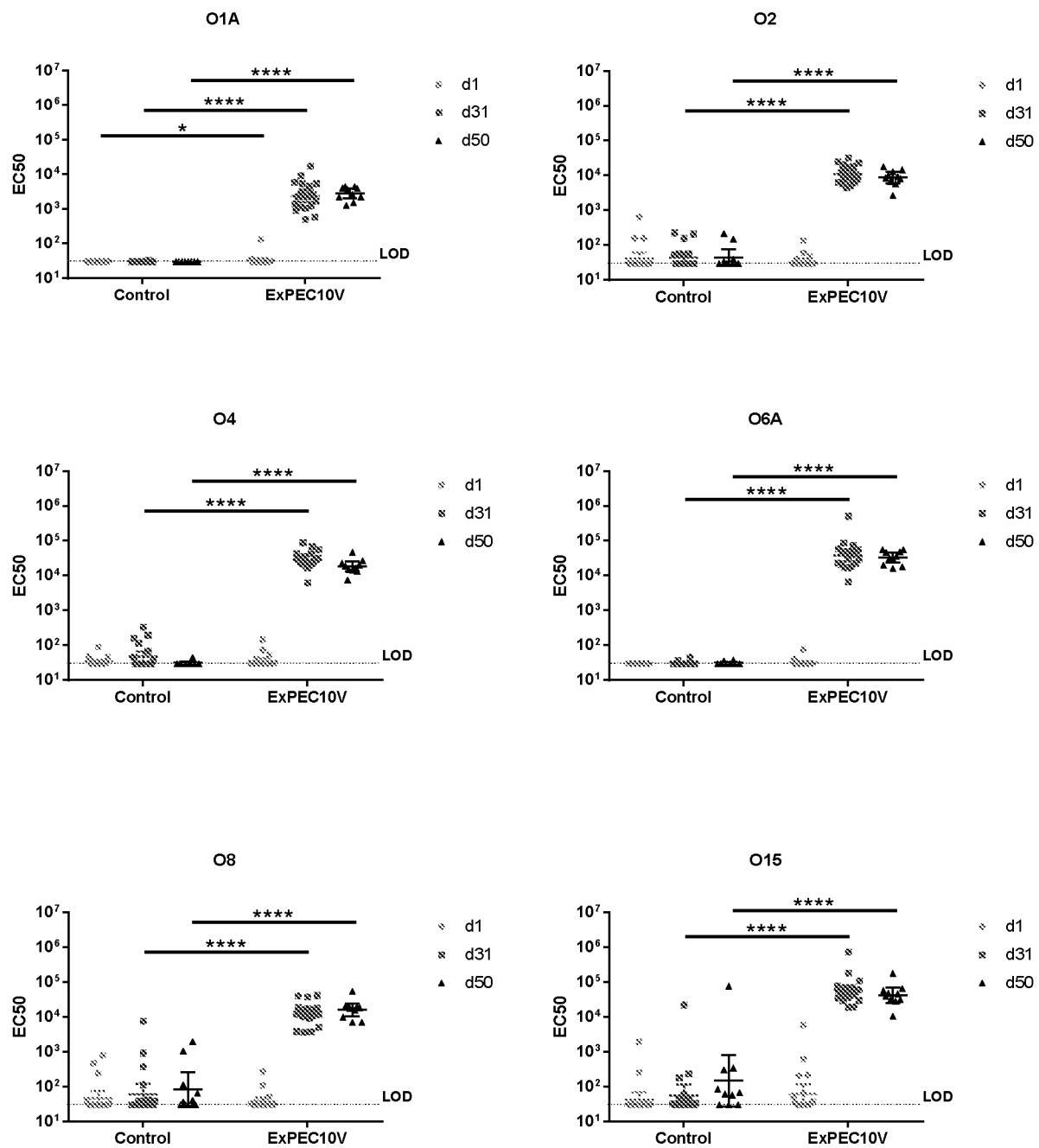


Fig. 9

10/12

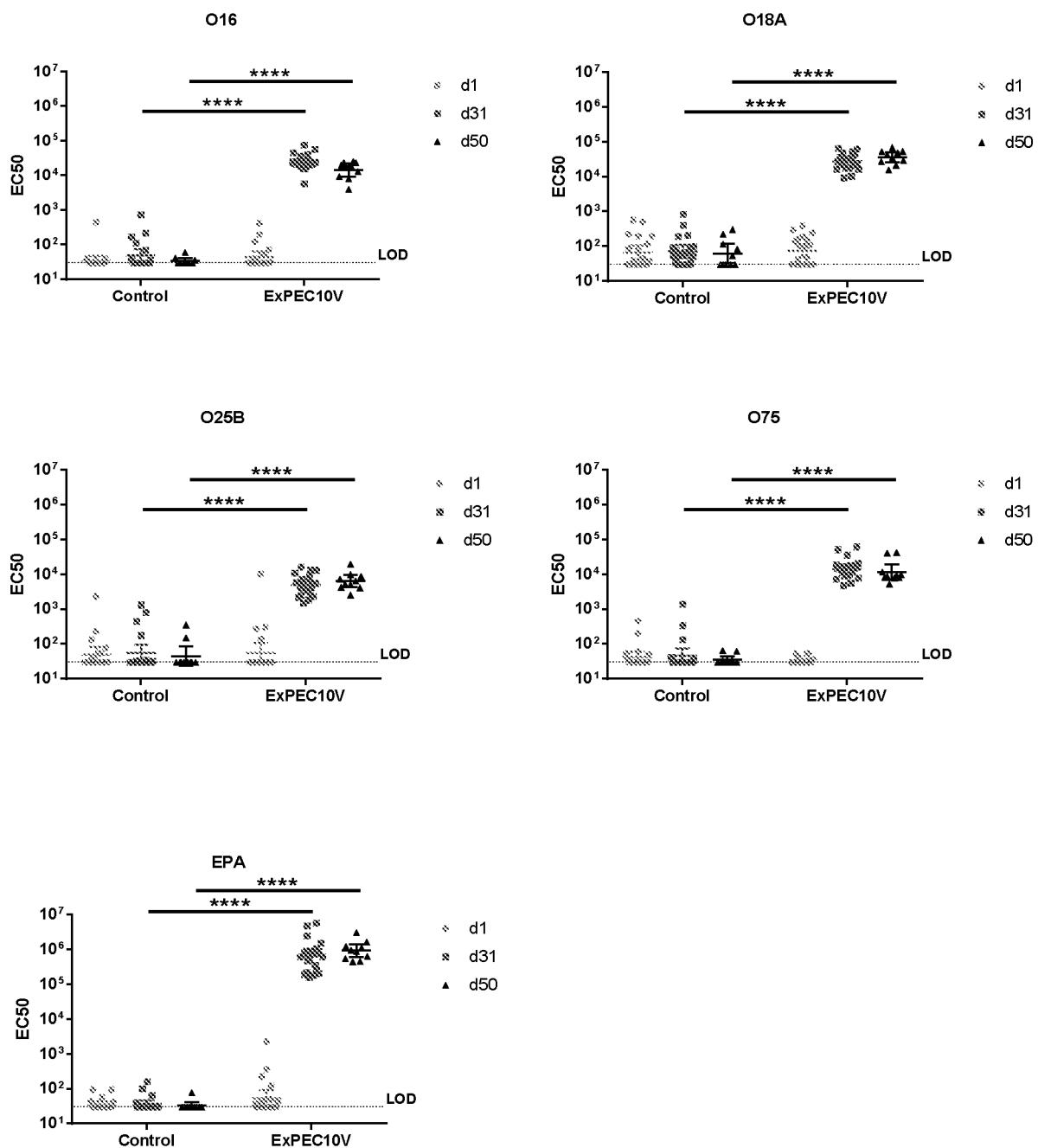


Fig. 9 - continued

11/12

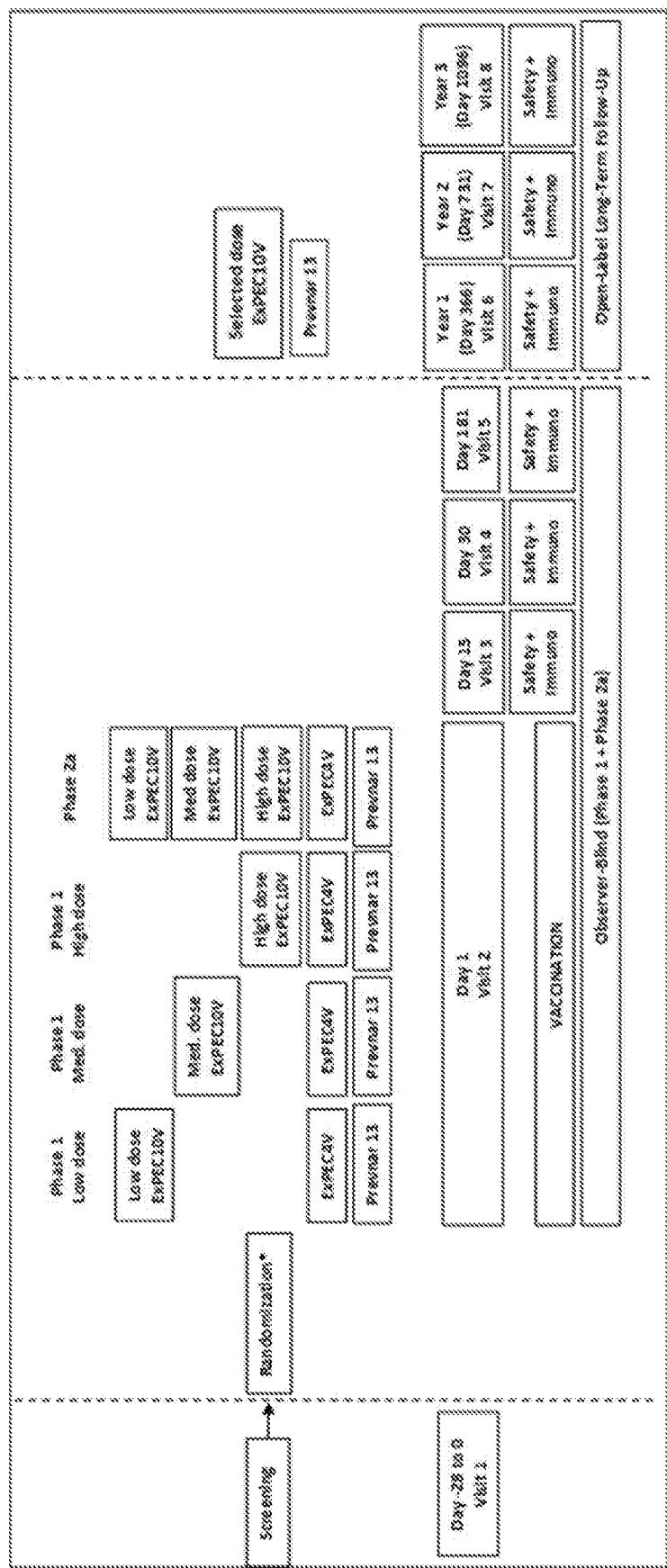
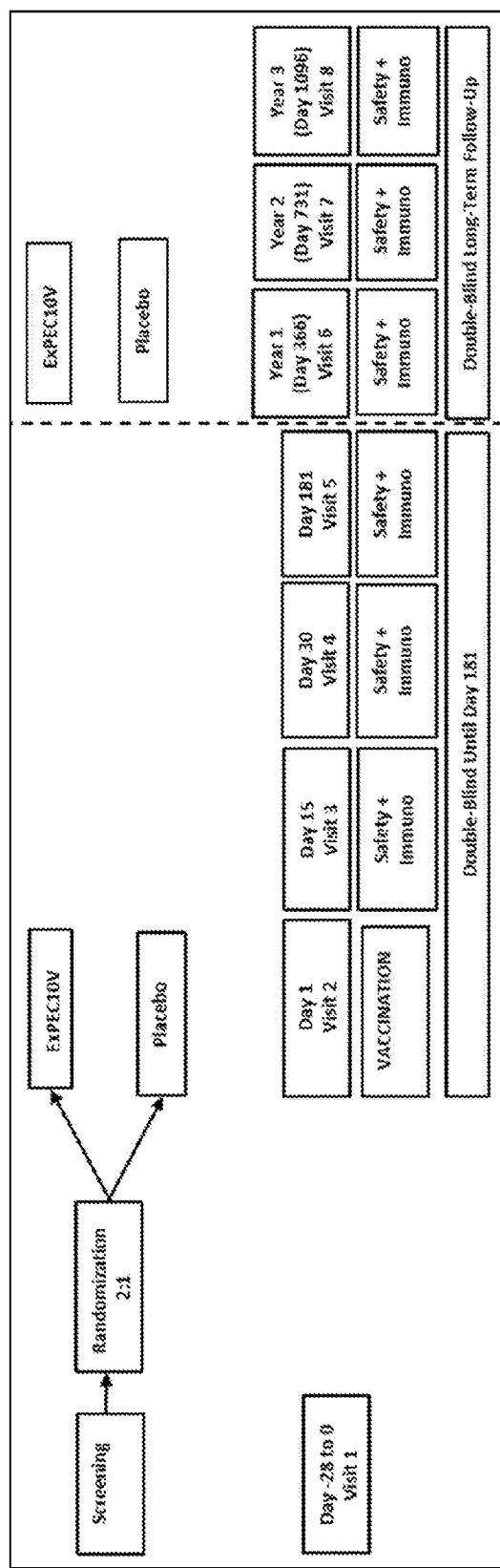


Fig. 10A

12/12

Fig. 10B



SEQUENCE LISTING

<110> Janssen Pharmaceuticals, Inc.
GlaxoSmithKline Biologicals S.A.

<120> Methods of Producing Bioconjugates of E. coli O-Antigen Polysaccharides,
Compositions Thereof, and Methods of Use Thereof

<130> 004852.11612/128W01 (CRU6009WOPCT1; VB66734P)

<150> US62/819,762

<151> 2019-03-18

<160> 19

<170> PatentIn version 3.5

<210> 1
<211> 3
<212> PRT
<213> Artificial Sequence

<220>
<223> glycosylation consensus sequence

<220>
<221> VARIANT
<222> (2)..(2)
<223> any amino acid residue except Pro

<220>
<221> VARIANT
<222> (3)..(3)
<223> Ser or Thr

<400> 1

Asn Xaa Xaa
1

<210> 2
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> optimized glycosylation consensus sequence

<220>
<221> VARIANT
<222> (1)..(1)
<223> Asp or Glu

<220>
<221> VARIANT
<222> (2)..(2)
<223> any amino acid residue except Pro

<220>
<221> VARIANT
<222> (4)..(4)
<223> any amino acid residue except Pro

<220>
<221> VARIANT
<222> (5)..(5)
<223> Ser or Thr

<400> 2

Xaa Xaa Asn Xaa Xaa
1 5

<210> 3
<211> 652
<212> PRT
<213> Artificial Sequence

<220>
<223> EPA carrier protein comprising 4 glycosylation consensus
sequences (EPA-4)

<400> 3

Gly Ser Gly Gly Asp Gln Asn Ala Thr Gly Ser Gly Gly Lys
1 5 10 15

Leu Ala Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys Ala Cys
20 25 30

Val Leu Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser Val Asp
35 40 45

Pro Ala Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr Ser Met
50 55 60

Val Leu Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala Ile Asp Asn Ala
65 70 75 80

Leu Ser Ile Thr Ser Asp Gly Leu Thr Ile Arg Leu Glu Gly Gly Val
85 90 95

Glu Pro Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gln Ala Arg Gly
100 105 110

Ser Trp Ser Leu Asn Trp Leu Val Pro Ile Gly His Glu Lys Pro Ser
115 120 125

Asn Ile Lys Val Phe Ile His Glu Leu Asn Ala Gly Asn Gln Leu Ser
130 135 140

His Met Ser Pro Ile Tyr Thr Ile Glu Met Gly Asp Glu Leu Leu Ala
145 150 155 160

Lys Leu Ala Arg Asp Ala Thr Phe Phe Val Arg Ala His Glu Ser Asn
165 170 175

Glu Met Gln Pro Thr Leu Ala Ile Ser His Ala Gly Val Ser Val Val
180 185 190

Met Ala Gln Ala Gln Pro Arg Arg Glu Lys Arg Trp Ser Glu Trp Ala
195 200 205

Ser Gly Lys Val Leu Cys Leu Leu Asp Pro Leu Asp Gly Val Tyr Asn
210 215 220

Tyr Leu Ala Gln Gln Arg Cys Asn Leu Asp Asp Thr Trp Glu Gly Lys
225 230 235 240

Ile Tyr Arg Val Leu Ala Gly Asn Pro Ala Lys His Asp Leu Asp Ile

245 250 255
Lys Asp Asn Asn Asn Ser Thr Pro Thr Val Ile Ser His Arg Leu His
260 265 270

Phe Pro Glu Gly Gly Ser Leu Ala Ala Leu Thr Ala His Gln Ala Cys
275 280 285

His Leu Pro Leu Glu Ala Phe Thr Arg His Arg Gln Pro Arg Gly Trp
290 295 300

Glu Gln Leu Glu Gln Cys Gly Tyr Pro Val Gln Arg Leu Val Ala Leu
305 310 315 320

Tyr Leu Ala Ala Arg Leu Ser Trp Asn Gln Val Asp Gln Val Ile Arg
325 330 335

Asn Ala Leu Ala Ser Pro Gly Ser Gly Gly Asp Leu Gly Glu Ala Ile
340 345 350

Arg Glu Gln Pro Glu Gln Ala Arg Leu Ala Leu Thr Leu Ala Ala Ala
355 360 365

Glu Ser Glu Arg Phe Val Arg Gln Gly Thr Gly Asn Asp Glu Ala Gly
370 375 380

Ala Ala Ser Ala Asp Val Val Ser Leu Thr Cys Pro Val Ala Lys Asp
385 390 395 400

Gln Asn Arg Thr Lys Gly Glu Cys Ala Gly Pro Ala Asp Ser Gly Asp
405 410 415

Ala Leu Leu Glu Arg Asn Tyr Pro Thr Gly Ala Glu Phe Leu Gly Asp
420 425 430

Gly Gly Asp Val Ser Phe Ser Thr Arg Gly Thr Gln Asn Trp Thr Val
435 440 445

Glu Arg Leu Leu Gln Ala His Arg Gln Leu Glu Glu Arg Gly Tyr Val
450 455 460

Phe Val Gly Tyr His Gly Thr Phe Leu Glu Ala Ala Gln Ser Ile Val
465 470 475 480

Phe Gly Gly Val Arg Ala Arg Ser Gln Asp Leu Asp Ala Ile Trp Arg
485 490 495

Gly Phe Tyr Ile Ala Gly Asp Pro Ala Leu Ala Tyr Gly Tyr Ala Gln
500 505 510

Asp Gln Glu Pro Asp Ala Arg Gly Arg Ile Arg Asn Gly Ala Leu Leu
515 520 525

Arg Val Tyr Val Pro Arg Trp Ser Leu Pro Gly Phe Tyr Arg Thr Gly
530 535 540

Leu Thr Leu Ala Ala Pro Glu Ala Ala Gly Glu Val Glu Arg Leu Ile
545 550 555 560

Gly His Pro Leu Pro Leu Arg Leu Asp Ala Ile Thr Gly Pro Glu Glu
565 570 575

Glu Gly Gly Arg Val Thr Ile Leu Gly Trp Pro Leu Ala Glu Arg Thr
580 585 590

Val Val Ile Pro Ser Ala Ile Pro Thr Asp Pro Arg Asn Val Gly Gly
595 600 605

Asp Leu Asp Pro Ser Ser Ile Pro Asp Lys Glu Gln Ala Ile Ser Ala
610 615 620

Leu Pro Asp Tyr Ala Ser Gln Pro Gly Lys Pro Pro Arg Glu Asp Leu
625 630 635 640

Lys Leu Gly Ser Gly Gly Asp Gln Asn Ala Thr

645

650

<210> 4
<211> 421
<212> PRT
<213> Artificial Sequence

<220>
<223> 04 GtrS amino acid sequence

<400> 4

Met Asn Asn Leu Ile Met Asn Asn Trp Cys Lys Leu Ser Ile Phe Ile
1 5 10 15

Ile Ala Phe Ile Leu Leu Trp Leu Arg Arg Pro Asp Ile Leu Thr Asn
20 25 30

Ala Gln Phe Trp Ala Glu Asp Ser Val Phe Trp Tyr Lys Asp Ala Tyr
35 40 45

Glu Asn Gly Phe Leu Ser Ser Leu Thr Thr Pro Arg Asn Gly Tyr Phe
50 55 60

Gln Thr Val Ser Thr Phe Ile Val Gly Leu Thr Ala Leu Leu Asn Pro
65 70 75 80

Asp Tyr Ala Pro Phe Val Ser Asn Phe Phe Gly Ile Met Ile Arg Ser
85 90 95

Val Ile Ile Trp Phe Leu Phe Thr Glu Arg Phe Asn Phe Leu Thr Leu
100 105 110

Thr Thr Arg Ile Phe Leu Ser Ile Tyr Phe Leu Cys Met Pro Gly Leu
115 120 125

Asp Glu Val His Ala Asn Ile Thr Asn Ala His Trp Tyr Leu Ser Leu
130 135 140

Tyr Val Ser Met Ile Leu Ile Ala Arg Asn Pro Ser Ser Lys Ser Trp

145

150

155

160

Arg Phe His Asp Ile Phe Phe Ile Leu Leu Ser Gly Leu Ser Gly Pro
165 170 175

Phe Ile Ile Phe Ile Leu Ala Ala Ser Cys Phe Lys Phe Ile Asn Asn
180 185 190

Cys Lys Asp His Ile Ser Val Arg Ser Phe Ile Asn Phe Tyr Leu Arg
195 200 205

Gln Pro Tyr Ala Leu Met Ile Val Cys Ala Leu Ile Gln Gly Thr Ser
210 215 220

Ile Ile Leu Thr Phe Asn Gly Thr Arg Ser Ser Ala Pro Leu Gly Phe
225 230 235 240

Ser Phe Asp Val Ile Ser Ser Ile Ile Ser Ser Asn Ile Phe Leu Phe
245 250 255

Thr Phe Val Pro Trp Asp Ile Ala Lys Ala Gly Trp Asp Asn Leu Leu
260 265 270

Leu Ser Tyr Phe Leu Ser Val Ser Ile Leu Ser Cys Ala Ala Phe Val
275 280 285

Phe Val Lys Gly Thr Trp Arg Met Lys Val Phe Ala Thr Leu Pro Leu
290 295 300

Leu Ile Ile Ile Phe Ser Met Ala Lys Pro Gln Leu Thr Asp Ser Ala
305 310 315 320

Pro Gln Leu Pro Thr Leu Ile Asn Gly Gln Gly Ser Arg Tyr Phe Val
325 330 335

Asn Ile His Ile Ala Ile Phe Ser Leu Leu Cys Val Tyr Leu Leu Glu
340 345 350

Cys Val Arg Gly Lys Val Ala Thr Leu Phe Ser Lys Ile Tyr Leu Thr
355 360 365

Ile Leu Leu Phe Val Met Gly Cys Leu Asn Phe Val Ile Thr Pro Leu
370 375 380

Pro Asn Met Asn Trp Arg Glu Gly Ala Thr Leu Ile Asn Asn Ala Lys
385 390 395 400

Thr Gly Asp Val Ile Ser Ile Gln Val Leu Pro Pro Gly Leu Thr Leu
405 410 415

Glu Leu Arg Lys Lys
420

<210> 5
<211> 1266
<212> DNA
<213> Artificial Sequence

<220>
<223> Example 04 gtrS nucleic acid sequence

<400> 5
atgaataatt taatttatgaa taactggtgt aaatttatcta tattttattat tgcatttatt 60
ttgctatggc ttagaaggcc ggatatactc acaaacgcac aattttgggc agaagattcc 120
gttttctggc ataaggacgc ctatgagaac ggattcttaa gttcaactaac aacgcctagg 180
aatgggtatt tccagactgt ttctacattt atagttggtc tgactgcttt attaaatcca 240
gattatgcac cttttgttgc taattttttt ggcataatga ttcgctcagt aatttatgg 300
tttttattta cagaaagatt caacttcctc acattgacta ctaggatttt cttatctatt 360
tattttctat gcatgcctgg attggatgaa gttcatgcaa atataacaaa tgcacattgg 420
tatttgtcat tatatgtatc aatgatcctg atagctcgca atccaagttc aaaatcatgg 480
aggtttcatg atatattctt tatcttgcta tccgggctca gtggcccatt tataatttc 540
attttagcag cttcatgctt taaatttata aataattgta aagatcatat tagtctaaga 600

tctttcataa	atttctactt	gcgtcagcca	tacgcattaa	tgattgtttg	cgcttaatt	660
caaggaacct	ctataattct	aactttcaat	ggcacacgtt	cctcagcacc	gctaggattc	720
agtttgcgt	tgatttcgtc	tattatatca	tcgaatattt	tttatattac	atttgtccca	780
tgggatattg	caaaggctgg	gtgggataat	ttactgttat	cttattttt	gtctgttcg	840
attttgcgt	gtgcggcctt	tgttttgtt	aaaggtacgt	ggcgaatgaa	agtatttgca	900
actttaccat	tgctaattat	aatatttca	atggcaaaac	cacaattgac	agactcggca	960
cctcaattgc	caacacttat	taatggcaa	ggttcaagat	acttcgtaaa	tatacatatt	1020
gcgatattct	cttgctatg	tgtttactta	cttgagtgcg	tcaggggaa	agtggcaact	1080
ttatTTCCA	aaatatactt	aacaattttg	ctattcgtga	tgggatgttt	gaattttgtt	1140
atcacccac	tcccaaacat	gaactggagg	gaaggtgcta	cttgattaa	taatgaaaa	1200
actggtgatg	tcatttcgat	tcaagtgcta	ccacctggcc	taacacttga	actaaggaaa	1260
aaataa						1266

<210> 6
<211> 713
<212> PRT
<213> Artificial Sequence

<220>
<223> Example PglB sequence ('wild-type')

<400> 6

Met Leu Lys Lys Glu Tyr Leu Lys Asn Pro Tyr Leu Val Leu Phe Ala
1 5 10 15

Met Ile Ile Leu Ala Tyr Val Phe Ser Val Phe Cys Arg Phe Tyr Trp
20 25 30

Val Trp Trp Ala Ser Glu Phe Asn Glu Tyr Phe Phe Asn Asn Gln Leu
35 40 45

Met Ile Ile Ser Asn Asp Gly Tyr Ala Phe Ala Glu Gly Ala Arg Asp
50 55 60

Met Ile Ala Gly Phe His Gln Pro Asn Asp Leu Ser Tyr Tyr Gly Ser
65 70 75 80

Ser Leu Ser Ala Leu Thr Tyr Trp Leu Tyr Lys Ile Thr Pro Phe Ser
85 90 95

Phe Glu Ser Ile Ile Leu Tyr Met Ser Thr Phe Leu Ser Ser Leu Val
100 105 110

Val Ile Pro Thr Ile Leu Leu Ala Asn Glu Tyr Lys Arg Pro Leu Met
115 120 125

Gly Phe Val Ala Ala Leu Leu Ala Ser Ile Ala Asn Ser Tyr Tyr Asn
130 135 140

Arg Thr Met Ser Gly Tyr Tyr Asp Thr Asp Met Leu Val Ile Val Leu
145 150 155 160

Pro Met Phe Ile Leu Phe Phe Met Val Arg Met Ile Leu Lys Lys Asp
165 170 175

Phe Phe Ser Leu Ile Ala Leu Pro Leu Phe Ile Gly Ile Tyr Leu Trp
180 185 190

Trp Tyr Pro Ser Ser Tyr Thr Leu Asn Val Ala Leu Ile Gly Leu Phe
195 200 205

Leu Ile Tyr Thr Leu Ile Phe His Arg Lys Glu Lys Ile Phe Tyr Ile
210 215 220

Ala Val Ile Leu Ser Ser Leu Thr Leu Ser Asn Ile Ala Trp Phe Tyr
225 230 235 240

Gln Ser Ala Ile Ile Val Ile Leu Phe Ala Leu Phe Ala Leu Glu Gln
245 250 255

Lys Arg Leu Asn Phe Met Ile Ile Gly Ile Leu Gly Ser Ala Thr Leu

260 265 270

Ile Phe Leu Ile Leu Ser Gly Gly Val Asp Pro Ile Leu Tyr Gln Leu
275 280 285

Lys Phe Tyr Ile Phe Arg Ser Asp Glu Ser Ala Asn Leu Thr Gln Gly
290 295 300

Phe Met Tyr Phe Asn Val Asn Gln Thr Ile Gln Glu Val Glu Asn Val
305 310 315 320

Asp Leu Ser Glu Phe Met Arg Arg Ile Ser Gly Ser Glu Ile Val Phe
325 330 335

Leu Phe Ser Leu Phe Gly Phe Val Trp Leu Leu Arg Lys His Lys Ser
340 345 350

Met Ile Met Ala Leu Pro Ile Leu Val Leu Gly Phe Leu Ala Leu Lys
355 360 365

Gly Gly Leu Arg Phe Thr Ile Tyr Ser Val Pro Val Met Ala Leu Gly
370 375 380

Phe Gly Phe Leu Leu Ser Glu Phe Lys Ala Ile Met Val Lys Lys Tyr
385 390 395 400

Ser Gln Leu Thr Ser Asn Val Cys Ile Val Phe Ala Thr Ile Leu Thr
405 410 415

Leu Ala Pro Val Phe Ile His Ile Tyr Asn Tyr Lys Ala Pro Thr Val
420 425 430

Phe Ser Gln Asn Glu Ala Ser Leu Leu Asn Gln Leu Lys Asn Ile Ala
435 440 445

Asn Arg Glu Asp Tyr Val Val Thr Trp Trp Asp Tyr Gly Tyr Pro Val
450 455 460

Arg Tyr Tyr Ser Asp Val Lys Thr Leu Val Asp Gly Gly Lys His Leu
465 470 475 480

Gly Lys Asp Asn Phe Phe Pro Ser Phe Ala Leu Ser Lys Asp Glu Gln
485 490 495

Ala Ala Ala Asn Met Ala Arg Leu Ser Val Glu Tyr Thr Glu Lys Ser
500 505 510

Phe Tyr Ala Pro Gln Asn Asp Ile Leu Lys Thr Asp Ile Leu Gln Ala
515 520 525

Met Met Lys Asp Tyr Asn Gln Ser Asn Val Asp Leu Phe Leu Ala Ser
530 535 540

Leu Ser Lys Pro Asp Phe Lys Ile Asp Thr Pro Lys Thr Arg Asp Ile
545 550 555 560

Tyr Leu Tyr Met Pro Ala Arg Met Ser Leu Ile Phe Ser Thr Val Ala
565 570 575

Ser Phe Ser Phe Ile Asn Leu Asp Thr Gly Val Leu Asp Lys Pro Phe
580 585 590

Thr Phe Ser Thr Ala Tyr Pro Leu Asp Val Lys Asn Gly Glu Ile Tyr
595 600 605

Leu Ser Asn Gly Val Val Leu Ser Asp Asp Phe Arg Ser Phe Lys Ile
610 615 620

Gly Asp Asn Val Val Ser Val Asn Ser Ile Val Glu Ile Asn Ser Ile
625 630 635 640

Lys Gln Gly Glu Tyr Lys Ile Thr Pro Ile Asp Asp Lys Ala Gln Phe
645 650 655

Tyr Ile Phe Tyr Leu Lys Asp Ser Ala Ile Pro Tyr Ala Gln Phe Ile

660

665

670

Leu Met Asp Lys Thr Met Phe Asn Ser Ala Tyr Val Gln Met Phe Phe
675 680 685

Leu Gly Asn Tyr Asp Lys Asn Leu Phe Asp Leu Val Ile Asn Ser Arg
690 695 700

Asp Ala Lys Val Phe Lys Leu Lys Ile
705 710

<210> 7

<211> 120

<212> PRT

<213> Escherichia coli

<400> 7

Met Leu Lys Leu Phe Ala Lys Tyr Thr Ser Ile Gly Val Leu Asn Thr
1 5 10 15

Leu Ile His Trp Val Val Phe Gly Val Cys Ile Tyr Val Ala His Thr
20 25 30

Asn Gln Ala Leu Ala Asn Phe Ala Gly Phe Val Val Ala Val Ser Phe
35 40 45

Ser Phe Phe Ala Asn Ala Lys Phe Thr Phe Lys Ala Ser Thr Thr Thr
50 55 60

Met Arg Tyr Met Leu Tyr Val Gly Phe Met Gly Thr Leu Ser Ala Thr
65 70 75 80

Val Gly Trp Ala Ala Asp Arg Cys Ala Leu Pro Pro Met Ile Thr Leu
85 90 95

Val Thr Phe Ser Ala Ile Ser Leu Val Cys Gly Phe Val Tyr Ser Lys
100 105 110

Phe Ile Val Phe Arg Asp Ala Lys
115 120

<210> 8
<211> 306
<212> PRT
<213> Escherichia coli

<400> 8

Met Lys Ile Ser Leu Val Val Pro Val Phe Asn Glu Glu Glu Ala Ile
1 5 10 15

Pro Ile Phe Tyr Lys Thr Val Arg Glu Phe Glu Glu Leu Lys Ser Tyr
20 25 30

Glu Val Glu Ile Val Phe Ile Asn Asp Gly Ser Lys Asp Ala Thr Glu
35 40 45

Ser Ile Ile Asn Ala Leu Ala Val Ser Asp Pro Leu Val Val Pro Leu
50 55 60

Ser Phe Thr Arg Asn Phe Gly Lys Glu Pro Ala Leu Phe Ala Gly Leu
65 70 75 80

Asp His Ala Thr Gly Asp Ala Ile Pro Ile Asp Val Asp Leu Gln
85 90 95

Asp Pro Ile Glu Val Ile Pro His Leu Ile Glu Lys Trp Gln Ala Gly
100 105 110

Ala Asp Met Val Leu Ala Lys Arg Ser Asp Arg Ser Thr Asp Gly Arg
115 120 125

Leu Lys Arg Lys Thr Ala Glu Trp Phe Tyr Lys Leu His Asn Lys Ile
130 135 140

Ser Asn Pro Lys Ile Glu Glu Asn Val Gly Asp Phe Arg Leu Met Ser
145 150 155 160

Arg Asp Val Val Glu Asn Ile Lys Leu Met Pro Glu Arg Asn Leu Phe
165 170 175

Met Lys Gly Ile Leu Ser Trp Val Gly Gly Lys Thr Asp Ile Val Glu
180 185 190

Tyr Val Arg Ala Glu Arg Ile Ala Gly Asp Thr Lys Phe Asn Gly Trp
195 200 205

Lys Leu Trp Asn Leu Ala Leu Glu Gly Ile Thr Ser Phe Ser Thr Phe
210 215 220

Pro Leu Arg Ile Trp Thr Tyr Ile Gly Leu Val Val Ala Ser Val Ala
225 230 235 240

Phe Ile Tyr Gly Ala Trp Met Ile Leu Asp Thr Ile Ile Phe Gly Asn
245 250 255

Ala Val Arg Gly Tyr Pro Ser Leu Leu Val Ser Ile Leu Phe Leu Gly
260 265 270

Gly Ile Gln Met Ile Gly Ile Gly Val Leu Gly Glu Tyr Ile Gly Arg
275 280 285

Thr Tyr Ile Glu Thr Lys Arg Pro Lys Tyr Ile Ile Lys Arg Val
290 295 300

Lys Lys
305

<210> 9
<211> 14440
<212> DNA
<213> Artificial Sequence

<220>
<223> example 04 rfb locus nucleotide sequence - 04-EPA production
strain BVEC-L-00684f

<400> 9
atgacgaatt taaaaggcagt tattcctgtat gcgggtctcg ggatgcataat gttgcctgcc 60
actaaggcga tacccaaaga gatgctacca atcgtcgaca agccaatgtat tcagtcatt
gttgacgaga ttgtggctgc agggatcaaa gaaatcctcc tgtaactca cgcgtccaag 120
aacgcggctcg aaaaccacctt cgacacctct tatgagtttag aatcactcct tgagcagcgc
gtgaagcgtc aactgctggc ggaagtacag tccatctgtc cggcggcggt gaccattatg 240
aacgtgcgtc agggcgaacc ttttagttta ggccactcca ttttgtgtgc ggcacccgtcc 300
attggtgaca acccatttgt cgtggtaactg ccagacgttg tgatcgacga tgccagcgcc
gaccgcgtac gttacaacct tgctgcatg attgcacgtt tcaacgaaac gggccgcagc 420
caggtgctgg caaaacgtat gccgggtgac ctctctgaat actccgtcat ccagactaaa 480
gagccgctgg accgtgaggg taaagtcaacg cgcattgttg aatttatcga aaaaccggat 540
cagccgcaga cgctggactc agacatcatg gccgttaggtc gctatgtgct ttctggcgat 600
atttggccgg aactggaacg tactcagcct ggtgcattgg gacgtattca gctgactgat 660
gctattgccc agctggcgaa aaaacaatcc gttgatgcaa tgctgatgac cggcgcacagt 720
tagcactgca gcaaaaaat gggctatatg caggcgtttg tgaagtatgg cctacgcaac 780
ctgaaagaag gggcgaagtt ccgtaaaggt attgagaagc tggtaagcga ataatgaaaa 840
tctgaccgga tgtaacggtt gataagaaaa ttataacggc agtggaaatt cgcagaaaa 900
gtaatttggcgtt gcgaatcttc ctggcgttgtt tttatataaa ccatcagaat aacaacgagt 960
tagcagtagg gttttattca aagtttcca ggatttcct tggttccaga gcggattgg 1020
aagacaatta gcgttgaat ttttcgggtt tagcgcgagt gggtaacgct cgtcacatca 1080
taggcattgtca tgcagtgtc tggtagctgt aaagccaggg gcggtagcgt gcattaaatc 1140
ctctattaaat caaaactgaga gccgcttatt tcacaggatg ctctgaagta atatgaaata 1200
aattaagtga aaatacttgt tactggtggc gcaggattta ttggttcagc tgttagttcgt 1260
cacattataa ataatacgca ggatagtgtt gttaatgtcg ataaattaac gtacgccgga 1320
aaccggaaat cacttgctga tgtttctgat tctgaacgct atgttttga acatgcggat 1380
atttgcgtatc cacctgcaat ggcacggatt tttgctcagc atcagccgga tgcagtgtatg 1440
1500

cacctggctg ctgaaagcca tggaccgt tcaattacag gccctgcggc atttattgaa 1560
accaatattg ttggtaactta tgtcctttg gaagccgctc gcaattactg gtctgcttt 1620
gatagcgaca agaaaaatag cttccgtttt catcatattt ctactgacga agtatatgg 1680
gatttgccctc atcctgacga ggttaataat acagaagaat tacccttatt tactgagaca 1740
acagcttacg cgccaaggcag cccttattcc gcatccaaag catccagcga tcatttagtc 1800
cgcgctgga aacgtaccta tggttaccg accattgtga ctaattgctc taacaattat 1860
ggtccttatac atttcccgga aaaattgatt ccattggta ttctcaatgc tctggaaggt 1920
aaagcattac ctattnatgg taaagggat caaattcgcg actggctgta tggtaagat 1980
catgcgcgtg cgttatatac cgtaacc gaaggtaaaag cgggtgaaac ttataacatt 2040
ggtgggcaca acgaaaagaa aaacatagat gtagtgctca ctattgtga tttgctggat 2100
gagattgtac cgaaagagaa atcttacgt gagcaaatac cttatgttgc cgatcgccg 2160
ggacacgatc gccgttatgc gattgatgct gagaatattt gtcgcaatt gggatggaaa 2220
ccacaggaaa cgaaaagagcg aaggattcgg aagacagtgg aatggtatct gtccaaatac 2280
aaatgggttataatgtgaa aagtgggcc tatcaatcgt ggattgaaga gaactatgag 2340
ggccgccagt aatgaatatac ctccttttgc gcaaaacagg gcaggttagt tgggaactac 2400
agcgtgctct ggcacctctg ggtaacttga ttgctttga tggttattcc actgattatt 2460
gtggcgattt cagtaacccc gaaggtgtgg ctgaaaccgt caaaaaaatt cgcccaatgt 2520
ttattgttaa tgctgctgct cataccgcgg tagataaggc tgagtcagaa ccagaatttg 2580
cacaattact caatgcgacc agcgttgaag caattgcaaa agcggctaattt gaagttgggg 2640
cttggtaat tcattactca actgactacg tctccctgg aaatggcgac atgccccatggc 2700
tcgagactga tgtaaccgct ccgctcaatg tttatggcaaa aaccaaattt gctggagaaa 2760
gagcattaca agaacattgc gcaaaagcatc ttatttccg taccagctgg gtatatgcag 2820
gtaaaggaaa taactttgcc aaaacaatgt tacgtctggc aaaagagcgc gaagaactgg 2880
ctgtgataaa cgatcagttt ggcgcacca caggtgctga attgctggct gattgcaccg 2940
ctcatgccat tcgcgtggca ttaaaaaaac cagaagttgc tggcttgtac catctggtag 3000

caa atggcac aacaacctgg cacgattacg ccgcgc tagt attcgaagaa gcccgtaaag 3060
cagggattga ccttgcactt aacaaactca acgccgtacc aacaacggct tatcctactc 3120
cagcccccg tcctcataat tctcgccctca ataccgaaaa gtttcagcag aactttgcgc 3180
ttgtcttgcc tgactggcag gtggcgtga aacgtatgct caacgaatta tttacgacta 3240
cgccaattta acaaattttt gcatctcgct catgatgccca gagcgggatg aattaaaagg 3300
aatggtgaaa taaaaacgcg taaaggtatt attctggctg gtggttccgg cactcgctt 3360
tatcctgtga cgatggcagt gagtaaaca ctgctgccga tttatgataa gccgatgatt 3420
tattatccgc tttcaacgct tatgttagcg ggtattcgcg atattcttat ttcagtgacg 3480
ccacaggata caccgcgttt ccaacaattt ttgggggacg ggagtcagtg ggggcttaat 3540
ctacagtata aagtacaacc gagtccggat ggcctggcgc aagcgtttat tattggtaa 3600
gactttattt gtggtgatga ttgtgcactc gtacttggcg ataatatctt ctatggacac 3660
gacttgcga aattaatgga agctgctgtt aacaaagaaa tcggtgcaac ggtatttgct 3720
tatcacgtca atgatcctga acgttatggt gtcgtggagt ttgataataa cggtaactgca 3780
attagcctgg aagaaaaacc gctggaacca aaaagtaact atgcggttac tgggctttat 3840
ttctatgaca atgatgttgt agaaatggcg aaaaacctta agcctctgc ccgtggcgaa 3900
ctggaaattt ccgatattaa ccgtattttt atggaggcagg gacgtttgc tgcgtatg 3960
atggggcgtg gttatgcctg gttggatact ggtacacatc aaagtcttat tgaagcaagt 4020
aacttcattt ccaccattga agagcgtcag ggattaaagg tatcttgcgg ggaagagatt 4080
gcttaccgta aagggtttat tgcgtgatc caggtgaaag tattagccga accgctgaag 4140
aaaaatgatt atggtcagta tctgctaaaa atgattaaag gttattaata aaatgaacgt 4200
aattaaaact gaaattcctg atgtgctgat ttttgaacca aaagttttt ggtatgaacg 4260
tggcttcttt tttgagagtt ttaaccagaa agtatttgaa gaagctgttag gacggaaggt 4320
tgaatttggtt caggataacc attctaagtc taaaataat gtattgcgtg ggatgcatta 4380
tcaaacacaa aatactcaag gaaaactggc tcggtaatt tctggttcag tatatgatgt 4440
tgccgttagat ttaagagaaa aatcaaagac atttggcaaa tgggtgggtg tagaattatc 4500

tgggaataat aaaagacaat tgtggatccc cgaaggaaaa tttatgtgtt 4560
ggaggagaat accgaatttg tttataaaatg taccgatact tataaccctg ctcatgaaca 4620
cacattgcta tggaatgatc caactatcaa tataagttgg ccaatcatac aaaactgcaa 4680
gccaaattatt tctgaaaaag atgctaattgg acatctttt tcacataaaa cctattctg 4740
aaatgcaata ttatgagttt aattagaaac agtttctata atattgctgg ttttgcgtgt 4800
ccgacattag ttgcagtccc tgcttgggg attcttgcca ggctgcttgg accggagaat 4860
tttggacttt tcacactagc attcgcttg ataggatatg caagtatttt cgacgccggg 4920
attagtcgag ctgtaatcgag agaaatcgct ctttatcgag aaagtgaaaaa agagcaaata 4980
caaattattt cgacagcaag tgtaatcgta ctattcttag gggtggttgc agcttgcgtt 5040
ctttatTTta gtagtaataa agttgttgag ttattgaatg ttagttccgt ttatattgaa 5100
acagcagtgc gtgcattctc tgTTTatttca tttataatac ctgtgtatct gattaaccag 5160
atttggcttg gttatctgga agggctagaa aaatttgcaa atataaatgt tcagagaatg 5220
atttcttagca caagcttggc tatattacca gtgatatttt gttattacaa tccctcgTTg 5280
ctttatgcta tgtatggTTt ggtggTTggg cgtgtgattt cattttgtat tagcgaata 5340
atttgcgag atattattct taaaagtaaa ctTTacttta atgtggcaac ttgcaatcgt 5400
cttatctctt ttggTggatg gataacagtt agtaatatca taagccccat catggcatat 5460
ttcgaccgct ttatcatctc tcatattatg ggggcttcga gaattgcatt ttatacagcg 5520
ccctcagagg gtgtatcaag gttattaaat atcccatatg cttggcaag agctctattt 5580
cctaaattgg catatagcaa taatgatgat gaacgaaaaa aattacaact acagagctac 5640
gcaattataa gcattgtatg tctacccata gttgttattt gttgtatTTt tgcctcattc 5700
ataatgacaa catggatggg acctgattat gccttagaag cagcaactat catgaaaata 5760
cttcttgctg gttttttctt taactcttta gcgcaaatac cttatgcata cttgcattct 5820
atcgaaagt caaaaattac cgcattgtg catctcatag aacttgcGCC atacttatta 5880
ttattgtatt acttcacaat gcatttcggc ataattggca cggcaatcgc ttggTCactt 5940
agaacatttt gtgattttgt tatactactt tcgatATcgA gaagaaaaatg attgcggTTg 6000

atattgcgt tgcaacctac aatggtgcta attttattcg gcaacagatt gaatctatcc 6060
agaaaacaac ttatagaaat tggcgtctta taataagtga tgataactcg agtgatgata 6120
ctgttgcata tattaaggat atgatgtcta acgacagtcg tatctatttg gtaggaaata 6180
aaagacaagg aggggttatt cagaacttta attatgctct ttcacaaact acatctgaaa 6240
tttgttact atgtgaccag gatgacattt ggccggagga gcgtctggaa attcttatag 6300
ataaatttaa ggccttgcag cgtaatgatt ttgttccggc aatgatgttt actgatttga 6360
aatttagtaga cggaaataat tgtttGattt cagaaagttt ttatcgaacg aataatatta 6420
atccacaaga taatctgaaa aataataatc ttctctggcg ttcaacggta tatggctgta 6480
cttgcatcat gaataagaaa cttgttgata ttgcattgcc tataacctaca tatgcacata 6540
tgcatgatca atgggtggca ttattagcga agcaatatgg taacatTTT tatttcgact 6600
atgcgtctgt tcgttatagg caacattcta caaatgttgt tggtggtaga aataaaacgc 6660
catttcaaaa atttaattcc atacaaaaaa acctaaaaag gattaatttg ctatggata 6720
gaactgttgc tttattttaa tcaaataacg atttctatcc agggaaataa atggaaaata 6780
aaattgatta cttaaaattt ggagtgaatg aagtattacc ttatctttt aaaggaaaca 6840
agaaagtttt ttcactttgt gtattaatta gtttggcatt acaaaaatga tatattttt 6900
atttttttt gcactgtta tgatctgtac gtttttaaca cacaggcgac aggattata 6960
tggttatct gcgttagtat ttctttttt ggcttttaacc tatccatcag gaggggactg 7020
gataggttat tttctccatt atgactgcat ggttaatgag cagtgtata atggttttat 7080
aatgtttgaa cctggatatg aattaattgt ttccttattt ggatatttg gatttcagac 7140
aattattatt tttatagccg ctgtaaatgt aattctaata ttaaattttt caaagcattt 7200
tgaaaacgga agttttgtta ttgttgcgt aatgtgcgt ttcctttgga gtgtttatgt 7260
tgaggcgatt agacaggctc tggccttac tatagttata tttgggattc attctttttt 7320
tttgggtaga aaaaggaaat ttataacatt agtattttt gcgtcaactt tccatataac 7380
tgctttgatt tgtttcttc taatgactcc tctatttca aagaaattaa gcaagataat 7440
aagttatagc ctattaattt tcagtagctt cttttcgct ttttctgaaa ccatattaag 7500

tgcaactcctt gcaatttgc cagaaggatc cattgccagt gaaaaattaa gttttactt 7560
agcaaccgag caatacaggc cacagttatc tattgggagt ggcactattc ttgacattat 7620
acttattttt ctgatatgtg taagtttaa acgaataaag aaatatatgc tcgctaatta 7680
taatgctgca aatgagatat tgcttattgg ttgctgtctt tatatttctt tcggatttt 7740
tatcggaaa atgatgccag ttatgactcg cattgggtgg tatggtttc cattgttat 7800
agtacttctt tatattaact tgggttattc agaatatttt aagaggtata taaataaaag 7860
agggtgtggg tatagcaaat tattaattgc tttttatttt ttgctacaaa ttttgcgacc 7920
attaacatat gattatagct attataatat aatgcaccag gatactttgc tgaataggtt 7980
tgatgcatta gatgatgcat cattaagaca atcagcgaag agaaaatgtt tcgatttggg 8040
aaagatagga tatggttct tatgttagtat ataatatcct gcattcattc ggataatttc 8100
ctatggaagt gtccttgct ctgtctgtcc tcatttgg 8160
aaattttatg ttaataagaa gctttagata accacttagg aactgtatgt ttgatctgtc caaaaattat attattgtaa 8220
gtgcgacggc gctggcttcc ggaggtgcat taactatatt aaagcaattt ataaaacatg 8280
catcacaaaa ttcaaatgac tatattatgt ttgtatctgc gggattggag ttgccggct 8340
gtgataacat catttacata gaaaacacac caaaaggatg gttgaaaaga atatattggg 8400
attggttcgg ttgtcggaag tttatctcgg aacataagat taacgttaag aaagtaattt 8460
ctctacaaaa ttccagtttgc aatgttcctt acgaacagat tatttacttg caccagccaa 8520
ttccttttag taaagttgat tccttttaa aaaatatcac atccgataac gtaaagctt 8580
ttttatataa aaagtttat tccttattta tatttaataa tgtgaatgcc aatacaacca 8640
tcgttgtc aacgaattgg atgaaaaaag gagtgctgga gcaatgtatgaaaatttagta 8700
ccgaaagggt ccttggataa aacactgata tcaaaggatt taataatact aattttgatg 8760
tagatatgga tgtatctgc aaaaacactct tatatccagc gacaccactt acctataaaa 8820
atcatttggt cattctgaag gcgttggtaa ttttaagaa aaagtatttt atagatgatc 8880
tgaaattcca agtgactttt gaaaagaata ggtacaaaaa ttttgcataag ttttgcaat 8940
taaataactt aagcaaaaac gttgattatc tcggcggttct ttcataactcg aacttgcaaa 9000

aaaaatata ggcggcatct ttaatcgaaa ttccatgacta tatcgaatca tatgggttac 9060
cactcatcga agctgctagt ttaggaaaaaa aaatcattag tagtgatctt ccttatgcc 9120
gggatgtttt aaaggattat agcggcgtag attttgaat ttacaataat gaagatggct 9180
gggctaaggc gttgttaat gttttaatg gcaattcgaa gctcaattttt aggccattatg 9240
aaaaagatag tcgttcatct tggccacagt tcttctctat tttgaataaa ggtgtattat 9300
gtttaatggc aaaatattgt taattactgg tggtacgggg tcttcggtt atgctgttct 9360
aagacgtttt cttgacactg atatcaaaga aatacgtatt tttcccggg atgaaaaaaaa 9420
acaagatgac atgagggaaaa aatataataa tccgaaactt aagttctata taggtgatgt 9480
tcgactat tcgagtatcc tcaatgcttc tcgagggtttt gattttattt atcatgctgc 9540
agctctgaag caagtacctt cctgcgaatt ccacccaatg gaagctgtaa aaacgaatgt 9600
tttaggtacg gaaaacgtac tggaagcggc aatagctaat ggagtttaggc gaattgtatg 9660
tttgagtaca gataaagctg tatatcctat caatgcaatg ggtatttcca aagcgatgat 9720
ggaaaaagta atggtagcaa aatcgcgcaaa tggtgactgc tctaaaacgg ttatttgcgg 9780
tacacgttat ggcaatgtaa tggcatctcg tggttcagtt atcccattat ttgtcgatct 9840
gattaaatca ggttagaccaa tgacgataac agaccctaat atgactcgat tcatgatgac 9900
tctcgaaagac gctgttgcattt tggttcttta cgcatggaa catggcaata atggtgat 9960
ttttgtccaa aaggcacctg cggctaccat cgaaacgttg gctattgcac tcaaagaatt 10020
acttaatgta aaccaacacc ctgtaaatat aatcgccacc cgacacgggg aaaaactgta 10080
cgaagcgat ttagccgag agggaaatgat tgcagccgag gatatgggtt attattatcg 10140
tggtccacca gatctcccgat atttgaacta tggaaaatat gtggaacatg gtgaccgtcg 10200
tatctcgaa gtggaagatt ataactctca taatactgat aggttagatg ttgaggaaat 10260
gaaaaaaatta ctgctaaaac ttccctttat ccgggcactt cggctgttg aagattatga 10320
gttggattca taatatgaaa attttagtta ctggcgctgc agggtttac ggtcgaattt 10380
tggttccg gcttaaggaa gctggatata acgaactcat tacgatagat cgtaactctt 10440
ctttggcgaa ttttagagcag ggacttaagc aggcagattt tattttcac cttgctgggg 10500

taaatcgtcc cgtgaaggag tgtgaatttg aagagggaaa tagtaatcta actcaacaga 10560
ttgttcatat cctgaaaaaa aacaataaaa atactcctat catgctgagt tcttcatcc 10620
aggctgaatg tgataacgct tatggaaaga gtaaagcagc tgccggaaaaa atcattcagc 10680
agtatgggaa aacgacaaac gctaaatatt atatttatcg ctgccgaat gtattcgta 10740
agtgggtcg accaaattat aactcctta tagcaacttt ctgccatcgc attgcaaatg 10800
atgaagctat tacaattaat gatccttcag cagttgtaaa tctggtgtat atagatgact 10860
tttggctga catattaaag ctattagaag gagcgaacga aactggttac aggacatttg 10920
gtccaattta ttctgttact gttggtaag tggcacaatt aatttaccgg tttaaagaaa 10980
gtcgccaaac attaatcacc gaagatgttag gtaatggatt tacacgtgca ttgtactcaa 11040
catggtaag ttacctgtct cctgaacagt ttgcgtatac ggcccttct tatagtgatg 11100
acagaggggt attctgtgaa gtattgaaaa cgaaaaacgc gggccagttt tcgttcttta 11160
ctgcgcattcc aggaattact cgggggtggtc attatcatca ttccaaaaat gagaattta 11220
ttgtcatccg aggaagtgc tgttcaat ttgaaaatat tgcacgagt gaacgatatg 11280
aacttaatgt ttcctctgat gattttaaa ttgttggaaac agttccgggaa tggacgcata 11340
acattactaa taatggctcg gatgagctag ttgttatgct ttgggcaaattt gaaatattta 11400
atcggtctga accagatact atagcgagag ttttatcgta aaaaaatttga aagtcatgtc 11460
ggttgttggg actcgccag aaattattcg actctcgctg gtccttgcaa aatttagatga 11520
atattgtgac caccttattt ttcataccgg gcaaaactac gattatgaac tgaatgaagt 11580
ttttttcaaa gatttgggtg ttgcggaaacc tgatttttt cttaatgccc caggtaaaaa 11640
tgcagcagag actattggac aagttatcat taaagttgat gaggtcccttg aacagggaaa 11700
accagaagcc atgttagtac ttggcgatac taactcctgt atttcagcaa taccagcaa 11760
gcgtcgaaga attccgatct tccatatgga ggctggaaat ctttggggat accaacgcgt 11820
accggaagaa actaacagaa aaatagttga tcataccgct gatataata tgacatata 11880
tgatatcgcg cgtgaatatac ttctggctga aggtgtacca gccgatagaa ttattaaaac 11940
cggttagccca atgtttgaag tactcactca ttatatgccc cagattgatg gttccgatgt 12000

actttctgc ctgaatttaa cacctggaa tttcttgc gtaagtgcc acagagaaga 12060
aaatgttcat acccctaaac aacttgtgaa actggcgaat atacttaata ccgtggctga 12120
aaaatatgtat gtcccggttag ttgtttctac tcatcctgc actcgtaacc gcatcaacga 12180
aaacggtatt caattccata aaaatatctt gcttcttaag ccattaggat ttcacgatta 12240
caaccatctg caaaaaaatg cacgtgctgt tttatcgat agtggacta ttacagaaga 12300
gtcctccatt atgaacttcc ctgcactcaa tatacgagaa gcgcacgaac gcccgaaagg 12360
cttcgaagaa ggggcagtaa tggatggcgg tcttgaatct gatgcgttt tacaggcatt 12420
agaaattatt gcaacacagc ctcgtggaga agtacgctta ctgcgtcagg ttagtgacta 12480
tagcatgcca aatgtttcag ataaagttct gcgtattatc cattcatata ctgactacgt 12540
taaacgggtt gtctggaagc aatactaataa aaacttgcata taatcattga tgattatttg 12600
ccccatagca cacgcgttgg ggctaaaatg tttcatgagt taggccttga attactgagc 12660
agaggccatg atgtaactgt aattacgcct gacatctcat tacaagcaat ttattctatt 12720
agtatgattt atggtataaa gggttggcgt ttcaaaatg gacctttaaa ggatgttaggt 12780
aaggctaaac gtgccataaa tgaaactctt ttatctttc gcgcattggcg cgcatatgg 12840
cacctcattt aacatgatac atttgcgtt atcggttttattt attccccctc tatttttgg 12900
ggcgacttgg ttaaaaaat aaaacaacga tgccagtgcc caagctatct gatcctaagg 12960
gatatgtttc cacagtgggt cattgatgca ggtatgttga aagccgggttc accaattgaa 13020
aaatatttttta ggtatgttga aaaaaatgtca tatcagcagg ctggccggat agggtaatg 13080
tctgataaga atcttgagat atttcgcctt accaataaaag gttatccgtt gtaagtttta 13140
cgtaattggg cctcaatgac tcctgtgtct gccagcgatg attatcatttctt acttcgtcaa 13200
aaatacgatc taaaagataa agtcattttt ttctatggcg gtaatattgg gcatgctcag 13260
gatatggcaa acttaatgcg ctttgcgtt aatatgtgc gttatcatga tgctcatttc 13320
ctgtttatag ggcagggtga tgaagttgag ctgataaaat ctcttgctgc agaatggaaat 13380
ttaactaattt tcactcatctt accttcgtt aaccagggaaag agttttaaattt aattttatct 13440
gaagttgttgc tcggcctgtt ctccctttca tctcgccattt cttcacataa tttccccggaa 13500

aaattactag ggtatatgg	tcaatcaatc ccgatcctg ggagtgtgaa tggcgcaat	13560
gatttaatgg atgtaattaa taagcacaga gccggtttca ttcatgttaa tggtgaagat		13620
gataaaactgt ttgaatctgc acaattgctt ctttgtatt cagtttaag aaaacagcta		13680
ggtcagaacg ctaatgtt gttaaagtct caatttcgg ttgaatcggc ggcacatact		13740
atcgaagtcc gactggaggc tggagaatgc gtttagttga tgacaatatt ctggatgaac		13800
tttttcgcac tgcagcaaat tctgaacgtt tgcgcgtca ttatatttgc cacgcacatctc		13860
atcaggagaa ggttcaacgt ttacttattt catttgtagt cgacagctat gttgaacccc		13920
attggcatga gttaccgcattt cagtggaaa tggttgcgt catgcaaggg caatttgcgtt		13980
tttgcgttgcgtt tgagcaaaat ggtgagatcc aaaaacagtt tggttgcgtt gacggtaggg		14040
gaataagcgt cgttgcgtt tccccaggag atatacatag tgtcaaattgc ctgtcaccaa		14100
aaggcccttat gttggagata aaggaggggc catttgaccc actcaaagct aaggcttttt		14160
ctaagtggtt atagggcgat acaccaccgt ttattcttct atcttattct atacatgctg		14220
ggttaccatc tttagcttctt caagccgcgc aacccgcgg tgaccacccc tgacaggagt		14280
agcttagcatt tgaccacccc tgacaggatt agcttagcata tgagctcgag gatatctact		14340
gtgggtaccc gggatccgtg taggctggag ctgcttcgaa gttcctatac tttcttagaga		14400
ataggaactt cggaatagga actaaggagg atattcatat		14440

<210> 10
 <211> 19
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> example signal sequence for EPA carrier protein

<400> 10

Met Lys Lys Ile Trp Leu Ala Leu Ala Gly Leu Val Leu Ala Phe Ser
 1 5 10 15

Ala Ser Ala

<210> 11
<211> 13043
<212> DNA
<213> Artificial Sequence

<220>
<223> example 01A rfb locus nucleotide sequence - 01A-EPA production
strain stGVXN4411 and stLMTB10217

<400> 11
atgacgaatt taaaaggcagt tattcctgtta gcgggtctcg ggatgcataat gttgcctgcc 60
actaaggcga tacccaaaga gatgctacca atcgtcgaca agccaatgtat tcagtcacatt 120
gttgcacgaga ttgtggctgc agggatcaaa gaaatcctcc tggtaactca cgcgtccaag 180
aacgcggctcg aaaaccacctt cgacacctct tatgagtttag aatcactcct tgagcagcgc 240
gtgaagcgtc aactgctggc ggaagtacag tccatctgtc cggccggcgt gaccattatg 300
aacgtgcgtc agggcgaacc ttttaggttta ggccactcca ttttgtgtgc gcgacctgcc 360
attggtgaca acccatttgt cgtggtaactg ccagacgttg tgatcgacga tgccagcgcc 420
gaccgcgtac gttacaacct tgctgccatg attgcacgtt tcaacgaaac gggccgcagc 480
caggtgctgg caaaacgtat gccgggtgac ctctctgaat actccgtcat ccagactaaa 540
gagccgctgg accgtgaggg taaagttagc cgcattgttg aatttatcga aaaaccggat 600
cagccgcaga cgctggactc agacatcatg gccgttaggtc gctatgtgct ttctgccat 660
atttggccgg aactggaacg tactcagcct ggtgcattgg gacgtattca gctgactgtat 720
gctattgccg agctggcgaa aaaacaatcc gttgatgcaa tgctgatgac cggcgcacagt 780
tacgactgctcg gcaaaaaaat gggctatatg caggcggttg tgaagtatgg cctacgcaac 840
ctgaaagaag gggcgaagtt ccgtaaaggt attgagaagc tgtaagcga ataatgaaaa 900
tctgaccgga tgtaacggtt gataagaaaa ttataacggc agtggaaatt cgcagaaaa 960
gtaatttggtt gcgaatcttc ctgccgttgt tttatataaa ccatcagaat aacaacgagt 1020
tagcagtagg gttttattca aagtttcca ggattttcct tgttccaga gcggattgg 1080
aagacaatta gcgtttgaat ttttcgggtt tagcgcgagt gggtaacgct cgtcacatca 1140

taggcatgca	tgcagtgc	tc	tggtagctgt	aaagccaggg	gcggtagcgt	gcattaatac	1200
ctctattaat	caaactgaga	gccg	cattt	tcacagcatg	ctctgaagta	atatggaata	1260
aattaagcta	gcgtgaagat	actt	gttact	aggggcgcag	gatttattgg	ttctgctgt	1320
gttcgtcaca	ttataaataa	tacgcaggat	agtgttgtt	atgtcgataa	attaacgtac	1380	
gccggaaacc	tggaatcact	tgctgatgtt	tctgactctg	aacgctatgt	ttttgaacat	1440	
gcggatattt	gcgatgctgc	tgcaatggcg	cggat	tttgc	ctcagcatca	gccggatgca	1500
gtgatgcacc	tggctgctga	aagccatgtg	gatcg	tttgc	ttacaggccc	tgcggcattt	1560
attgaaacca	atattgttgg	tacttatgtc	cttttggaaag	cggctcgcaa	ttactggtct	1620	
gctcttgatg	gcgacaagaa	aaatagcttc	cgtttc	atatttctac	tgacgaagtc	1680	
tatggtgatt	tgcctcatcc	tgacgaagta	aataataaag	aacaattacc	cctctttact	1740	
gagacgacag	cttacgcgcc	tagtagtcct	tattccgcat	caaaagcatc	cagcgatcat	1800	
ttagtccg	cgtggaaacg	tacctatgg	ttaccgacta	ttgtgactaa	ctgttcgaat	1860	
aactacgg	tc	ttccggaaaa	ttgattccac	tagtaattct	taatgctctg	1920	
gaaggtaagg	cattacctat	ttatggcaaa	ggggatcaaa	ttcgtgactg	gctgtatgtt	1980	
gaagatcatg	cgcgtgc	tttgcgtt	atataccgta	gttactgaag	gtcaagcggg	tggaaacctat	2040
aacattggcg	gacacaacga	aaagaaaaac	atcgatgtt	tgctgactat	ttgtgatttgc	2100	
ttggacgaga	tagtccc	gaa	agagaaatct	tatcg	tgact	tgtgtgtat	2160
cggccagg	atgatcgcc	ttatgc	gatt	gtgactgaga	agattgg	cgaaattgg	2220
tggaaaccac	aggaaacgtt	tgagagtgg	attcgtaaaa	cgg	tggaatg	gtattgg	2280
aatgc	aaaat	gggtt	gataa	tgt	aaaagt	gtgcctatc	2340
tatgagg	ggcc	gc	atgc	aatatc	ctcc	tttttgg	2400
aactacagcg	tgctctgg	ca	cctctgg	attt	gattgc	tcttgatgtt	2460
attactgtgg	tgat	tttagt	aacc	ctgaag	gtgtgg	ctga	2520
ctgatgttat	tgttaat	gtc	gcgg	ctcaca	ccgc	agtaga	2580
aatttgcaca	attactcaat	g	gcgact	actagcg	ttgaatcaat	tgcaaaagcg	2640
						gcaaatgaag	

ttggggcttg ggtaattcat tactcaactg actacgtatt ccctggaaat ggccacacgc	2700
catggctgga gatggatgca accgcaccgc taaatgttta cggtaaacc aagtttagctg	2760
gagaaaaagc attacaagag cattgtgcga agcaccta at tttccgtacc agctgggtct	2820
atgcaggtaa aggaaataat ttgc cccaaa cgatgttgcg tctggcaaaa gagcgtgaag	2880
aactagccgt tattaatgtat cagtttggtg cgccaaacagg tgctgaactg ctggctgatt	2940
gtacggcaca tgccattcgt gtcgcactga ataaaccgga tgtgcagggc ttgtaccatt	3000
tggtagccag tggtaccaca acctggtacg attatgctgc gctggttttt gaagaggcgc	3060
gcaatgcagg cattcctt gcactcaaca agctcaacgc agtaccaaca actgcctatc	3120
ctacaccaggc tcgtcgtcca cataactctc gccttaatac agaaaaattt cagcagaatt	3180
ttgcgcttgtt attgcctgac tggcagggtt gttgtaaacg catgctcaac gaattattta	3240
cgactacagc aatttaatag ttttgcatc ttgttcgtga tggtggagca agatgaatta	3300
aaaggaatga tgaaatgaaa acgcgtaaag gtattattttt agcgggtgg tctggactc	3360
gtcttatcc tgtgactatg gtcgtcagta aacagctattt acctatatata gataaaccga	3420
tgatctatta tccgctttctt acactgatgt tagcgggtat tcgcgatattt ctgattattta	3480
gtacgccaca ggataactcctt cgtttcaac aactgctggg tgacggtagc cagtggggcc	3540
tgaatcttca gtacaaagtgc caaccgagtc cggatggtct tgccgcaggca tttattatcg	3600
gtgaagagtt tattgggtggt gatgattgtg ctttggact tggtgataat atcttctacg	3660
gtcacgacctt gcctaaggta atggatgccg ctgttaacaa agaaagtggt gcaacggat	3720
ttgcctatca cgttaatgtat cctgaacgct atgggtcgat tgagtttgcg aaaaacggta	3780
cggcgatcag cctggaagaa aaaccgctac aaccaaaaag taattatgcg gtaaccggc	3840
tttattttta tgataacgac gttgtcgaaa tggcgaaaaa tcttaagcct tctgcccgcg	3900
gtgaactgga aattaccgat attaaccgta tctatatgga acaagggcgt ttatctgttg	3960
ccatgatggg gcgtggttat gcgtggtagt acacggggac acatcagagc ctgattgggg	4020
caagcaactt tattgcaaca attgaagagc gtcagggctt gaaagttcc tgcccgaaag	4080
aaattgctta ccgtaaaggg tttgttgcgatc ctgagcaggt gaaagtattt gctgaacctc	4140

tgaaaaaaaaaa tgcttatggc cagtatctgc taaaaatgtat taaagggttat taataaaatg 4200
aacgttaatta aaacagaaaat tcctgatgtc ctgatttttgc aaccgaaagt ttttgggtat 4260
gagcgtggtt tcttttttgc gagcttaac cagaagggtt ttgaggaagc tgttaggccgc 4320
aaagttgaat ttgttcagga taaccattcg aagtctagta aaggtgtttt acgcgggctg 4380
cattatcagt tggaaccta tgcacaagga aaattgggtgc gttgcgttgt cggtgaagtt 4440
tttgacgtat ctgttgatat tcgtaaatcg tcatcgactt ttggcaaatg gggtgggtg 4500
aatttatctg ctgagaataa gcggcaatttgc tggattcctg agggatttgc acatggttt 4560
ttagtgctga gtgagacggc ggagtttttgc tataagacga caaatttata tcatcctcag 4620
agtgatagag gaataaaaatg ggatgatcca agcatcaata tttcatggcc agtcgattca 4680
caagtgcgc tatcagctaa agataataag catcctccat taacaaagat taaaatgtat 4740
agttaagatc acgataaaatc ttggaaagggt tgcaaaatttgc aataaaaatag tgagcaaaag 4800
tgaataagg aacgtaatcc acaatgctgg ctatatgtat attactcaga tagctttata 4860
tggcacca ttatttatac tgagttatct gttaaaaaca ctgggggttg cacagtttgc 4920
taattatgcc ttaataactat caatcggtgc atatttacag attataacgg attatggttt 4980
ttcttttagt gcaagtcgtg cgatctcaca gaatagagag gacaaagaat atatatcaaa 5040
aatttatctg tcaactatgtc ctatcaagtt ggcgatatgc gctttcttat tcttattgct 5100
catgctattt ttaaatctt tgcctgtgc agctgaatta aaacaaggaa tattatatgg 5160
atatcttctt gtaataggaa atacttcca accacaatgg ttttccaag gtatcgaaaa 5220
ataaaaatc atagccctt ctaatgttat atcaagatgc gccgcgtgtt tacttgtatt 5280
tatctatgtg aggaatagcg aggatttaca aaaagcactt ttagtacagt cacttccatt 5340
agtaatttct gcgattggat taaatataatt tatattgaaa tatataata ttattttcc 5400
ggaaaaaaaaa ttatttaagg taattttaaa agaaggtaag gatTTTTTC ttgcataact 5460
ttattctgtt attctcaata atagtggcat ttttcttatttgc gggattttta ctaatcctgt 5520
tattgttgtt gtatatgccg ccgctgaaaa gatagtcaag gccgtattgt cgctatttac 5580
accactgacg caagctataat atccttataa ttgtcgtaag ttttcaactat ccgtatttgc 5640

cggcattgag gcagcaaaaa aaactggtat accaattata attttagcat ttatagctgc	5700
tgttatcggt gcaattaccc tacctgttgc aatcgactat cttatTTTC caaaagaaac	5760
aatTTTgtA ggtcaaatat taagtgcatt gatTTTTT ggtgttctta ataatgtatt	5820
cggcattcag atattgagtg catcaggaag aagtaaaata tatagttagga tggtattcgt	5880
atcagcgctt ataacattac ttttgattac tctattattt cagTTTgtA acgccactgg	5940
agtggcatgt gcaatattat tgggtgaaat gttcttatca atattgttac ttaaggata	6000
taaaaaaaa atttaaggaa tagttatgaa gaagttatta ttagtgttcg gtactaggcc	6060
tgaagcaata aagatggcct ctatcattga attattaaaa aaagattgtA gattcgaata	6120
taaaatatgt gtgacaggcc aacataaaga gatgcttgat caagttatgc aagtatttga	6180
tgttaaacct gattataatt tacggattat gcagcctggg caaacattag tatctatagc	6240
aacaaatata ctctcacggt taagtgaagt tttaattata gaaaagccag atattatact	6300
tgtgcattggg gatacaacga ctacccttgc tgctacttta gctgggtatt accaccaat	6360
aaaagtttgt catgtggaag caggattaag aacaggggat atttactctc cttggcctga	6420
agagggcaat cgtaaagtta cagggcatt agcatgtatt catttcgccc caacagagag	6480
atcaaaagat aatctcctga gggaggggt caaagtaaat aatataatttA taacggtaa	6540
taccgtcatc gactctttat ttattgcaaa agatatcata gataatgacc ctaatataaa	6600
gaacgcttta cataataat ttaattttct tgataaaagc cgacgagtag tacttataac	6660
aggtcatcga agagaaaatt tcgggaaagg ttttgaagat atatgcttg caataaagga	6720
attagcttc atttaccta atgttagattt tatttattccg gtgcattta atccaatgt	6780
aatggaacca gtacatcgta tattagataa tatagttaat attacctta ttgagccctt	6840
ggattatttA cttttgttt atttaatgaa tgagtcatat ttaatattga ctgattcagg	6900
ggggatacaa gaagaagcgc cttcgtagg taaaccggtt ttggttatgc gtgataactac	6960
tgaacccct gaggcggttg aggctggtaC tggtgtttaA gtggggactt ctaagataaa	7020
aatagtaaat aaagtaacgg agctattaaa caatgctgat atctacaatg ctatgtctc	7080
gttacataat ccatatggcg atgaaacagc tgctcaaaaa attcttaatg tgctcgcccA	7140

agagctaatt taatttaagc taaaaatatg ttattaatta ttgctgatta tccaaacgaa	7200
atgaatatgc gcgagggagc tatgcaacga atagatgcga tagactctct cattcgagat	7260
cgcaagcgag tgtatggaa tatttcattc aaaaaggcatc tagttcgctc aaatagttcc	7320
ttaataatg ttatagttga aaatctaaat gcaattattc acagaaacat cataaaacag	7380
tacatgcaaa aatcaacaac tatatatgtt cattctgttt ataatttatt aaaggttata	7440
acgctcattg atctaaaaaa aacaattctt gatatacatg gtgttgtacc ggaagaactt	7500
ttggcagata ataaaaaatt acttagtaaa gtatataaca tggtggaaaa aaaaggtgtc	7560
cttggatgca aaaaattaat acacgtcagt acagaaatgc aaaaacacta tgaagcaaaa	7620
tatggagtaa acttggctga aaggtaata gtgctcccga ttttgaata taaaatata	7680
acccaatcgc aaaacaaatg gacagaaaat aaaatacgaa gtatctatct tggaggatta	7740
caaacatggc aaaatattga taaaatgatt caagtttgc atgacacagt gataaacaat	7800
gaagcaggtt agtatgaatt caacttttc atcccacaga gtaacttgga agggtttata	7860
gataaatatt cgtaaaattt acataatatc aatgctaattt catctacgct atcacgtgat	7920
gaagtaattc cctttctaaa agaatgtcat attgggttttgc tattgcgcga tgatataata	7980
gtaaacagag ttgcgtgcc tacaaaattt gttgaatatt tagagtgtgg tgtcgttcca	8040
gttgtgctct ccccacttat aggtgattttt tattcgatgg gatataataa cattactaca	8100
gaggaaatgg ctaacagaag tataagtttgc ttggatcttgc aaaaatggc tgcacataat	8160
ttacaaattt tgacttctta tcagaagaga acctacaagg cacagaaaga acttattgct	8220
caactgtgct gaattttta catatataaa attatgttgc catatgcgg gtcaggtaat	8280
tgtatgcgtt tcaaataataa agataacggt tatataattt gtttctattt atgtttcatt	8340
ttgagctact tagtttactt caaatctgac tactttccttgc ctgattttctt gccatataca	8400
gaaatatacg atgggacata cgagaaatc aataatatttgc agcctgcctt tttatatttgc	8460
acacgggttgc ttcattttttt aaatttcccc tatataattttt ttgcaatgtt agtttgc	8520
ttatgtttaa gttggaaaat aaaatatgca agaaaaataa ttaaagatag ttatataat	8580
ttgttcttgc atgtatgtt atcatttttgc atgaaatgac tcaattgcgc	8640

atagcaattg cagtcactat gtgctatgtc tcggtttattt attactttta taaaaattgt	8700
attaaacatg cactgccatg gatgggtttg gctattttgtt ttcattacag cgcccttgctt	8760
ttatttatgtt cattatttat atacagttat aggaggttat taatagtaat tatagggttt	8820
gtaatatgtt tgagctttt aaacgtgtat gcagatacaa ttgcactata tttgccaat	8880
gaaaaaatag taaattttt atatagtattt tcattcatcat tagacaatag aaatgatttg	8940
gcaatattca acctgaataa tataatattt ttatcaatat ttatggat ctttatctt	9000
agccgatata taaaattaaa tgataatgag gcgaagttt ttaagtatgt gcaatgttca	9060
ggaatatttag cctttgtat tttcttctg gctagtggag tcccggtcat tgcttacgat	9120
actgcagagt tgctgcgaat atttatccg atggcttag tattaatcct ttgcata	9180
aaaaataata atatgcgtt ttttattgca gtcattatag ttatccttc aggcttaatg	9240
ttgtttataa cactaaggc tgtatcaata gttggtaag gattataaaa tgaatgttgc	9300
tatTTTGTG tctacgtata atggcggaaa atatTTAGAG gaacaactgg attcattgct	9360
gcttcaaagt ttcaggatt ttgttagtgc tatccgtat gacggatcat ctgatagaac	9420
tgtaaatata ataaaccaat acgtaatgaa agataacaga tttattaacg tggtaattc	9480
agaaaaatctt ggTTGTGCTG cttcgTTTAT taatttatta agaaatgctt cagccgat	9540
ttatATGTTT TGTGACCAAG atgattatttgc gcttccgaat aaattacagc gtgcgttgaa	9600
ttatTTTGCTG gctattgatc ctttacaacc tacattgtat cattgcgtatc taagcgttgt	9660
tgatgaaaaa cttaatatta tacaaaatttcc atTTTGCAG catcagaaaaa tgtcagcgta	9720
tgattcaatg agaaaaaata atctttcat acaaaattttt gttgttggtt gttcatgtgc	9780
tgttaatgct tcacttgcgg aatttggctt ttcgcgaattt ggagagcagc atgtaaaaat	9840
gatagctatg catgactggt ggTTGCCGT gactgcggggg cttttgggtt gaatccattt	9900
tgataatact caaacgatttcc tttatcgaca acatcaggc aatgtatttgc gtgcggaaatc	9960
atcaggtatg atgcgttttta ttgcatttgcg attaaatggg caagggattt cgcgagtttt	10020
atcttttgcg aaaaaagttt gttgcggaaa taagcttctt ttagatgtct atgataaaga	10080
tttaaatctt gagcaaaaaa aatctatcag gcttgcattt gaggcctt aagagaactc	10140

ttcaattgct gacctttaa aatgttctta tcatggtagc tatatgcaag gttttaaacg 10200
taatcttgcc ttaatatatt cagttctta cacaaaaaaa agaagatagt gtatccttat 10260
gaaaaaaaaatt gctattatcg gtactgttgg cataccagca tcataatggcg gatttgaac 10320
attagttgaa aatttaacaa gatacaattc ctcggagtt gaatataatg tttttgttc 10380
atcgttcac tacaaatccc accaaaaaaa acataatggg gcccgttaa tttatattcc 10440
gcttaaagcc aatggatggc agagcattgc gtatgacata atttcgttag catattctat 10500
tttttgaag cctgatgtga ttctgatttt aggggtttct ggttgttcat ttttgccttt 10560
cttcaaactc ttaacacgcg ctaagttat tactaatatt gatggcctgg aatggcgaag 10620
agataaatgg aattcaaaag tgaaacgttt cttaaaattt tcagaaaaaa tcgcagttca 10680
atattcgat gtcgttatta cgataatga ggcaatttct gagtacgtt ttaacgagta 10740
taataaagat agccgagta ttgcctatgg aggggatcat gcatggtaa atactgagga 10800
tgtatttaca acaagaaatt ataaaagcga ttactacctt tctgtatgtc gtatcgaacc 10860
cgaaaaacaat gtagaattaa ttttaaaac attttcaaag ctaaaatata aaataaaatt 10920
tattggaaat tggaatggca gcgagtttgg aaagaaactt aggctgcatt attctaacta 10980
tccaaatatt gaaatgattt atccgattta tgatcttcaa caattatttc acttacgaaa 11040
taattgcata ggatatatac atggcatttc ggctggagga acaaaccctt cttagtcga 11100
ggcaatgcat tttagtaaac ctatattgc atatgattgt aagtttaata ggtacactac 11160
tgaaaatgaa gcatgttatt tttctaatga atctgacctc gcagagaaaa tcataatgca 11220
ttgtgagcta tcattaggtg tctctggcac gaaaatgaaa gaaattgcta accagaata 11280
cacttggaga cgaatagcag aaatgtatga ggattgctat taactctgtt aaacttcaaa 11340
tcttttacaa tataatggcat gactataagc gcattaattt tttttcaagc cgctctcg 11400
gtgaccaccc cctgacaggg gatccgtgtaa ggctggagct gcttcgaagt tcctataactt 11460
tctagagaat aggaacttcg gaataggaac taaggaggat attcatatgg ataaagccgt 11520
aagcatataa gcatggataa gctatttata cttaataag tactttgtat acttatttgc 11580
gaacattcca ggccgcgagc attcagcgcg gtgatcacac ctgacaggag tatgtaatgt 11640

ccaagcaaca gatcggcgta gtcggtatgg cagtcatgg acgcaacctt gcgcctaaca	11700
tcgaaagccg tggttatacc gtctctattt tcaaccgttc ccgtgagaag acggaagaag	11760
tgattgccga aaatccaggc aagaaactgg ttccctacta tacggtaaa gagtttgtcg	11820
aatctctgga aacgcctcgt cgcatcctgt taatggtaa agcaggtgca ggcacggatg	11880
ctgctattga ttccctcaaa ccatatctcg ataaaggaga catcatcatt gatggtgta	11940
acacccctt ccaggacact attcgtcgta atcgtgagct ttccagcagag ggcttaact	12000
tcatcggtac cgggtttct ggcggtaag agggggcgct gaaaggcct tctattatgc	12060
ctggtgccca gaaagaagcc tatgaattgg tagcaccat cctgacaaa atgcgcgcg	12120
tagctgaaga cggtaacca tgcgttacct atattggtgc cgatggcgca ggtcactatg	12180
tgaagatggt tcacaacggt attgaatacg gcgatatgca gctgattgct gaagcctatt	12240
ctctgcttaa aggtggcctg aacccacca acgaagaact ggcgcagacc tttaccgagt	12300
ggaataacgg tgaactgagc agttacctga tcgacatcac caaagatatc ttcaccaaaa	12360
aagatgaaga cggtaactac ctgggtgatg tgatcctgga tgaagcggct aacaaaggta	12420
ccggtaatg gaccagccag agcgcgtgg atctcggcga accgctgtcg ctgattaccg	12480
agtctgtgtt tgcacgttat atctttctc taaaagatca gcgtgttgcc gcatctaaag	12540
ttctctctgg tccgcaagca cagccagcag gcgacaaggc ttagttcattc gaaaaagttc	12600
gtcgtgcgt gtatctggc aaaatcgttt cttacgccc gggcttctct cagctgcgtg	12660
ctgcgtctga agagtacaac tggatctga actacggcga aatcgcaag atttccgtg	12720
ctggctgcat catccgtcgc cagttcctgc agaaaatcac cgatgcttat gccgaaaatc	12780
cacagatcgc taacctgttg ctggctccgt acttcaagca aattgccat gactaccagc	12840
aggcgctgcg ttagtgcgtt gcttatgcag tacagaacgg tattccgggtt ccgacccct	12900
ccgcagcggc tgcctattac gacagctacc gtgctgtgt tctgcctcgc aacctgatcc	12960
aggcacagcg tgactatttt ggtgcgcata cttataagcg tatcgataaa gaagggtgt	13020
tccataccga atggctggat taa	13043

<211> 13790
<212> DNA
<213> Artificial Sequence

<220>
<223> example 02 rfb locus nucleotide sequence - O2-EPA production
strain stGVXN4906

<400> 12
atgacgaatt taaaaggagt tattcctgtta gcgggtctcg ggatgcataat gttgcctgcc 60
actaaggcga tacccaaaga gatgctacca atcgctgaca agccaatgtat tcagtagcatt 120
gttgcacgaga ttgtggctgc agggatcaaa gaaatcctcc tggttaactca cgcgtccaag 180
aacgcggctcg aaaaccactt cgacacctct tatgagtttag aatcactcct tgagcagcgc 240
gtgaagcgtc aactgctggc ggaagtacag tccatctgtc cggccggcgt gaccattatg 300
aacgtgcgtc agggcgaacc ttttaggttta ggccactcca ttttgtgtgc gcgacactgcc 360
attggtgaca acccatttgt cgtggtaactg ccagacgttg tgatgcacga tgccagcgc 420
gaccgcctac gttacaacct tgctgccatg attgcacgtt tcaacgaaac gggccgcagc 480
caggtgctgg caaaacgtat gccgggtgac ctctctgaat actccgtcat ccagactaaa 540
gagccgctgg accgtgaggg taaagtcaacgc cgcattgtt aatttatcga aaaaccggat 600
cagccgcaga cgctggactc agacatcatg gccgttaggtc gctatgtgct ttctggcgat 660
atttggccgg aactggaacg tactcagcct ggtgcattgg gacgtattca gctgactgat 720
gctattgccc agctggcgaa aaaacaatcc gttgatgcaa tgctgatgac cggcgcacagt 780
tacgactgacg gcaaaaaaaat gggctatatg caggcggtt tgaagtatgg cctacgcaac 840
ctgaaagaag gggcgaagtt ccgtaaaggt attgagaagc tggtaagcga ataatgaaaa 900
tctgaccgga tgtaacggtt gataagaaaa ttataacggc agtggaaatt cgcagcaaaa 960
gtaatttggtt gcgaatcttc ctggcggtt tttatataaa ccatcagaat aacaacgagt 1020
tagcagtagg gttttattca aagtttcca ggattttcct tggccaga gggattgg 1080
aagacaatta gcgttgaat tttcgggtt tagcgcgagt gggtaacgct cgtcacatca 1140
taggcatgca tgcagtgcctc tggtagctgt aaagccaggg gggtagcgt gcattaaatc 1200
ctctattaat caaactgaga gccgcttatt tcacagcatg ctctgaagta atatggaata 1260

aattaagtga aaataacttgt tactggtggc gcaggattta ttggttcagc tgttagttcg 1320
cacattataa ataatacgca ggatagtgtt gttaatgtcg ataaattaac gtacgccgga 1380
aacccggaaat cacttgctga tgtttctgat tctgaacgct atgaaaaatggat 1440
atttgcgatg cacctgcaat ggcacggatt tttgctcagc atcagccgga tgcagtgatg 1500
cacctggctg ctgaaagcca tggaccgt tcaattacag gccctgcggc atttattgaa 1560
accaatattg ttggacttta tgccttttgc gaagccgctc gcaattactg gtctgcttt 1620
gatagcgaca agaaaaatag cttccgtttt catcatattt ctactgacga agtctatgg 1680
gatttgcctc atccagatga agtaaataat acagaagaat tacccttatt tactgagacg 1740
acagcttacg cgccaaggcag cccttattcc gcatccaaag catccagcga tcatttagtc 1800
cgcgcatgga aacgtacgta tggtttaccg accattgtga ctaattgctc gaacaactat 1860
ggtccgtatc acttcccggaa aaagcttatt ccattggta ttcttaatgc actggaaagg 1920
aaggcattac ctattnatgg caaagggat caaattcgcg actgggtgttga tgtagaggat 1980
catgctcgtg cgttatatac cgtcgtaacc gaaggtaaag cgggtgaaac ttataacatt 2040
ggcggacaca acgaaaagaa aaacatcgat gttgtgtga ctattnatgg 2100
gagattgtac cgaaagagaa atcttacgt gagcaaatta cttatgttgc tgcgtcccc 2160
gggcatgatc gcccgttatgc aattgatgcc gataaaatta gcccgaatt gggctggaaa 2220
ccacagggaaa cggttggagag cgggattcgc aaaacggtgg aatggtatct ggctaataca 2280
aattgggttg agaatgtgaa aagggtgtct tatcagtcat ggatcgaaca aaactatgag 2340
ggccgtcagt aatgaatatc ctgttttcg gcaaaacagg gcagggtgggt tgggaactgc 2400
agcgtgctct ggcgcgcgtg ggtaatctga tcgcttttgc tggactcc actaattatt 2460
gtggagattt cagcaacccc gaagggtgtgg cagaaaccgt caaaaaatt cgtcctgacg 2520
ttattgttaa tgctgctgttgc cacactgcag tagataaagc agaatcagaa ccggatttcg 2580
cacaattact taacgcgaca agcgtcgaaag cgattgcaaa agctgctaattt gaaatcgcccc 2640
cctgggttat acactactct actgattatg tttccagg cagtggtgac ggcgcgtggc 2700
tgaaaaacggaa tgcaacagca ccgctaaatg tttacgggtga aacaaaattt gctggggaaa 2760

aggcattaca agaacattgc gcaaagcatc ttatTTCCG taccagctgg gtatacgctg 2820
gtAAAGGAAA taactttgct aaaacgatgt tgcgttggc aaaagaacgc gaagaactgg 2880
ctgtgataaa cgatcagttt ggCGCACCAA caggtgctga attgctggct gattgcaccg 2940
ctcatGCCat tcgcgtggca ttAAAAAAAC cagaagtgc tggcttgtac catctggtag 3000
caagtggcac aacaacctgg cacgattatg ctgcgctggt tttgaagag gcgcgcaaag 3060
cagggattaa tcttgcaCTT aacAAACTTA acGCCGTGCC aacaacggcc tatcccacac 3120
cagcccgtcg accccataac tctcgCCTCA atacagaaaa gttcagcag aactttgcgc 3180
ttgtcttgcc tgactggcag gtggcgtga aacgtatgct caacgaatta tttacgacta 3240
cggaattta acaaattttt gcatctcgct catgatGCCA gagcgggatg aattaaaagg 3300
aatggtggaaa tgaaaacgcg taaaggtatt attctggctg gtggttccgg cactcgtctt 3360
tatCCTGTGA cgatggcagt gagtaaaca ttgctGCCGA tttatgataa gCcgtatgatt 3420
tattatCCGC tttcaacgct tatgttagcg ggtattcgcg atattcttat tattagtacg 3480
ccacaggata caccgcgtt ccaacaatta ttggggacg ggagccagtg gggcttaat 3540
ctacagtata aagtacaacc gagtccggat ggcctggcgc aagcgtttat tattggcga 3600
gactttattg gtggtgatga ttgtgcactc gtacttggcg ataatatctt ctatggacac 3660
gacttGCCGA aattgatgga agctgctgtt aacaaagaaa gCGGTGCAAC ggtatttgct 3720
tatCACGTta atgatcctga acgtatggt gtcgtggagt ttgataataa cggtaCGGCA 3780
attAGCCTGG aagaaaaacc gctggagCCA aaaagcaact atgcggttac tgggctttat 3840
ttctatgaca atgacgttgt ggaaatggct aaaaaccta agccttctgc ccgtggcga 3900
ctggaaattta ccgatattaa ccgtatTTAT atggaacaag gacgtttgtc tgtagccatg 3960
atggggcgtg gctatgcATG gttggatACA gggacgcATC aaAGCCTTAT tgaAGCAAGT 4020
aacttcattg caacaattGA agagcgtcAG ggattAAAGG tatCTTGCCTT GGAAGAGATT 4080
gcttaccgta aagggtttat tgatGCCGAG caggtgaaAG tattAGCCGA accgcttATC 4140
aagaatcaat atggtaata tttgctgaaa atgatcagcg aatagtataat gggAACTCAA 4200
tgatggatAT taaattaATC tctttgcaAA aacatgggGA tgagcgcggT gcattaattG 4260

ctcttgaaga gcaacgaaat ataccttcg aagtcaaaag aatatattac atacttgaga 4320
ctcttaatgg agtaagacgc ggatttcatg cgcacaagg tactcgtcag ttagctattg 4380
tagtcaaggg agcttgtaaa tttcatctgg ataatggtaa agaaacaaag caggtggaac 4440
ttaatgatcc aacaattgcg ttgctgatag aaccctataat atggcatgaa atgtatgatt 4500
ttagtatgta ttgtgtgctg cttgtattt cgatgattt ctataaagag tctgattata 4560
tccgcaatta ttagtatttt attagaagag taaattcaat tgagaattca taagctaagt 4620
gacgtccaga caacatcaat tggatgatgaa acaactatct ggcagttgt tgtgatacta 4680
aaagggtgctg taattggtaa taattgcaac atctgtcaa atacctaatt tgaaaataac 4740
gttgtattt gtaacaatgt cacagtcaaa agcggtgtgt atatttggaa tggcgtaaaa 4800
atagaggata atgttttat tggccttgt gtacattta caaatgataa gtatcctcgc 4860
tctaaagtct atcctgatga attttgcaa acaataatac gcaaaggagc atcaataggt 4920
gctaacgcaa ccatcctgcc aggaattgaa attggtaaaa aagcaatcg tggcgcccc 4980
agtgttgtaa caaaaatgt accgcccattgc gcaatagtag taggtatcc agctcgattt 5040
attaaatggg tagaggataa tgaataaaat tgatttttt gatcttttcaatccca 5100
gcgacagcac aaagaattag tctctgcgtt tagtagggtg ctagattctg gttggatat 5160
catggcgaa gaacttgagc agttcgagaa agagttcgca gaatactgtg gagttaaat 5220
ttgcatttgtt gtacaaatgt gccttgatgc gttgatacta gtattgaggg catggaaaga 5280
acttggctat cttgaagacg gtgacgaggtt attagtaccg gcaaatacat atattgcttc 5340
tattcttgct ataacagaga acaaacttgt tcctgttctt gtgtaccat atatgaaac 5400
ttataatatt aatcctgctt taattgaaaa ttacattacg gaaaaacta aagcaatatt 5460
accggttcac ttatatggtc tattgtcaa tatgccagaa attagtgcac tcgcccggaaa 5520
atataatctg ttgattctt aagattgtgc acaagcacat ggtgcataac gtgtatggcgt 5580
caaagctgga gcttgggggg atgctgcagg atttagttt tatccaggaa aaaacccctgg 5640
agctttgggg gatgcggggag ctgttactac aaataatgca gaattatcct caactataaa 5700
agctttgcga aattatgggt cacataagaa atatgaaaat atttatcagg gattgaatag 5760

tcgattggat gaactgcaag cagccttatt gcgtgtaaaa atccatacat taccggaaga 5820
tactgcgatt cgcaaaggaa ttgctgaaaa atatattcgt gaaataaaaa accctgcgat 5880
tacgttacca gtgtacgaag gccaaagggtgc gcatgtttgg catttatttg tagtaagaat 5940
cgctaattcgt gaaaaattcc agtcatactt attagagaag ggtatcaaaa ccttaattca 6000
ctatccatta ccacccata agcagcaagc atatcaaaat atgtctagcc ttagccttcc 6060
aattactgag caaattcatg atgaagtcat ttcttacct ataagtccgg taatgagtga 6120
agatgatgtc aattatgtaa tcaaaatggt caatgattac aagtaatgaa aaaatttctt 6180
caggttaacta tattatccgc tatctataca ttcattaaaa tgattgcggg ttttatcatc 6240
ggtaaggtag tagcaattta tacagggcca tcagggtag caatgcttgg ccaagtgcaa 6300
agtttaatca caatagttgc aggtactacc tctgcacctg taagcacagg cttgttcga 6360
tatactgcgg aaaattggca agaaggacaa gaagcatgcg cgccatggtg ggcgcgtgc 6420
ttaagggta ctctgtttt attcttgctt attattcccg ttgttattat attgtcgaaa 6480
aatatttagtg agttactttt tagcgatgga caatacacat ggttaatcat tttcgcatgt 6540
tgtatattgc cattctccat tataaataca ttgatcgctt cagttttaaa tggtaaccaa 6600
ttttataagc aatatatatt gggtggatg ttttctgtat tcatttctac tatgtttatg 6660
attttgtga ttgttagctt taatcttaaa ggtgcattga ttgccacagc tataaatagt 6720
gctattgctg gtcttgatt ggttttattt tgtctcaata aatcttggtt tagatttaaa 6780
tattggtggg gtaaaacgga taaagacaaa attataaaaa ttattcatta tactctgatg 6840
gctctggttt ctgttatctc catgcctaca gcattgatgt gtattagaaa aatattgatt 6900
gctaaaactg gttgggagga tgcagggcaa tggcaggccg tatggaagat atctgaggtt 6960
tatcttggtg ttgtgacaat tgcttgcata acatatttct taccaagatt gacaattata 7020
aaaacaagtt tccttataaa aaaagaagta aatagtacta tattatacat aatatctatt 7080
acttcattca tggcggttag tatctattta ttccgcgatt tgtaataac agtttattt 7140
actgaacagt ttgcgtcagc tcgtgaatta ttttattac aacttatagg ggatgtata 7200
aaaattgctg ggttcttta tgcataaccct cttcaaagtc agggcatac taaactattc 7260

atcagttcag aagtgattt ttctatgctc tttatcatta ccacctata ttttgttga 7320
aattatggag tacatggtgc taacataagt tatgtcatta catatagttt atattttgtg 7380
tttgcattt tgtttactaa ttttattaat gttagaagaa ataattaaaa acagaggttg 7440
aattttgaaa ataattatac ctgtcttagg atttggcagg gctggtggtg aaagagttct 7500
ttctaagctg gcaactgaat tcatgttattt tggacatgtat gtaagtttg ttgttccaga 7560
taatagaact aatccatatt atgctaccac agcaaaaatt gtcacgagta aatctagtca 7620
aaaccgtgta aaaatattga gaatcattaa aaattactat aatctgtggc gttaatgcatt 7680
agagttaaat cctgatgctg tagttgcttag tttcatttgc actgcctatc ttgtcgcatt 7740
attaccaatc acccgctgta agaaatatta ttatattcag gcgtatgaag ttaatttttt 7800
tgataatata atatggaaat taatagcggg tttaacatata tatttaccgc ttaaaaaaaaat 7860
actaaatagt cctaatttgc ttcctcataa acatgtatgtat tttataggag tagttcctgc 7920
aggagtagat tttaaacgttt tctatccgaa accatcaaata agtttattaa atggcacac 7980
atcaataggg attattggta gaaaagagaa gcacaaagga actagcgaaa ttatttcagt 8040
attgtgttca ctggaaaata aagctggaat tataatcaat attgcgatct atcttgaaga 8100
agttgataag cagcgtttaa tcgctgccgg gtttcagggtt aatttttttc cgattacttc 8160
tgatttagaa ttggcatcct tttatcgaag caatgacatc atgattgctg ttgggttaat 8220
tgaagatggc gctttccatt atccttgc tgaatcaatg gcttgggtt gtcttggat 8280
ttcaaattat gcgccactta ctgaaactaa cagtgtactt aaattagtca agtttgatgc 8340
ttgcaaactt ggtgaagcaa ttaatctttg tctcaatctt gacctagaag aaaaaagcaa 8400
agaaaatccaa tctaataatctt ctgtgttggaa taaatatgac tggaaaatttgggtgaaac 8460
tttcaatagt ttattgttag atgcaaataa atagtatacg ttgatggggaa aatatgaat 8520
attgttaaaa ctgatattcc agatctgatc gttcttgaac caaaagtgtt tagtgtatgaa 8580
cgccggctttt ttatggagag ttataatcag attgaatttg agaaggcaat aggaaggcac 8640
gtaaaattttg ttcaaggataa tcattcaaaa tctagtaaag gcgtactacg tgggttgcatt 8700
tatcaattag caccgtatgc acaggctaaa ttagttcgat gtgttggtagg tcaggtatctt 8760

gatgttgctg ttgatcttag aaaaaattca ccaacgttca aaaaatggtt tggaaataacc 8820
ctttccgcag aaaataaacg acaattatgg atacccgaag gatttgctca tggttcttg 8880
gtgaccagtg atgaagctga gttcatttat aagacaacta actactatgc tcctggcat 8940
cagcaagcaa ttatcacaa tgatcctatt ttaaacatcg attggcctt ctgcagtagt 9000
gctctgtcat tacacaaaa agatcaagaa gcaaaattat tttcagaatt attggacagt 9060
gaactgttct aataaagtgt gccacctt ccgtctgaag gataggtggt tgcttatatt 9120
ttttttagta tgtttgtata atgacagaaa atagtcgaa atataaacac gataaaagct 9180
taataagttt tatctactta tttttatatt ttacacttatt tgtaggctt attatcgaa 9240
ataccagtt tttggggcga agtagagact atgataatta tatacagatc ttttctggta 9300
aagaagggga gggggttctt gaattatattt atcgcggatt gatgttaata acgaccagct 9360
atgaaactat cattttata attttaacat gttttttt tataaaggca aggtttctcg 9420
ctaactattc gcgtaattt tcaggcttga ctttattctt tatttattat gcaagcgtt 9480
cacttgggt ttttagattt actcaattca gaaatggtct atgtatttcc attttaatgt 9540
tttccgtata ctatattt ataaataaac cgacttattt ttatttctcg gtattatgt 9600
caattgcaac tcattggtct gctttgcctt ttttgcctt atatcctttt gtctattcaa 9660
caaaaataag acgccttgggt tattttgtt tcagtattct tttttgatt gcgatctcag 9720
gagaaggaaa agagatcata tctttataa gaaattttgg agtgggacaa aaaataggaa 9780
atgaagctgg tgtaaatttata ataaattcat tatcccttac cgctatttcc tggtttattt 9840
ttagttacat atcaaggattt gaaatgaaa ggagaaattt aaggctttc ttttgtttag 9900
gtgtcatgca atacgtgact tttagcctt tctctctacc ttttatggct ttccgtattt 9960
tggaaatgtt tttttcctt atgctaaccat ttgggggttt tattaagcaa aaaaagaattt 10020
attatatttattt ttttgcaaa gtgttaattt tattgtatct aacatactat tatcatatgg 10080
tctttggagt gattaatgtg taaggcttac gtgtggcta taattgttac ttacaacccg 10140
gaaatttttc gattgacgga atgtatttac tcttttagccc cacaagttga gagaataattt 10200
cttggtagata atggctcaaa taatagtgtat ttgataaaaa atatcagtat taataacctt 10260

gaaatttatt tacttcgga aaacaaaggc attgcattt ctcagaacca tgggttaag 10320
aagggcctgg aagcaaaga gtttgactat ttatTTTct cagatcagga tacttgctt 10380
ccttagcgatg ttattgaaaa acttaagagt acatttacga aaaataataa aaaaggtaaa 10440
aatgttgctt gtgcttctcc ttttttaaa gaccatcggtt caaattatat gcatccgtca 10500
gtcagcctaa atatTTTAC gagtacaaaa gttatATGTA gtgaagttaga cgatgatctt 10560
tatccctcgc atgttattgc ttctggatg ttaatgtctc gtgaagcatg gcgcgtcg 10620
ggaccatttt gtgaaaaact ctttataGAC tgggttgata cagaatggtg ttggcgtgca 10680
ttAGCTAATA atatgattat tgTTcagaca ccatcagtca tcatttctca tgaacttggg 10740
tatggcaga aaattttgc tggTCGATCT gttacaatac ataattctt cagaaatttt 10800
tataaaatac gcaatgcaat atacttaatg ctgcattcaa attatagctt caagtatcgt 10860
tatcatgctt ttttcatgc gacaaagaat gttgtatttG aaattttata ttcgaaagaa 10920
aaattaaatt cactgaaggt ttgtttaaa gctgtacgtg atggtagttt caataatttt 10980
taatacgaaa atagtttaggc tcaaggtgtt taaatggaag aaaataataat gaagacggc 11040
gctgttagttg gcacagtggg tgTTcctgct tgTTatggtg gttcgaatc acttggtag 11100
aatctaatttG attatcaatc tGatggtata caatatcaga tattttgctc ttcaaaaaaa 11160
tatgataaaaa aattttaaaa ttataaaaat gcagaattaa tctatTTGCC gataaatGCC 11220
aatggcgtct ctagcataat ttatgatatt atgtgtttaa ttatttggttt attcaaaagg 11280
ccagatgttG tttaatatt ggggtgtct ggttgtttat ttctaccaat ttataaacta 11340
ttttcaaaat caaagattat tgtcaatatt gatggcTTG aatggcgttag aaataaatgg 11400
ggaacgtttG ctaagaaatt tctaaaata tctgaggcga tatctattag aatagctgat 11460
attatcattt cagataatca agcaatagct gattatgtgg aaaataagta caagaaaaaa 11520
agtgttagtta tagcttatgg cggagatcat gccactaatc tttagtacacc gatagacaat 11580
gatcaaaaaa aagaaggta ttatttgggg ctttgttagga tagagcctga gaataatata 11640
gaaatgattc tgaatgcctt catTAataca gataaaaaaa ttaaattttat ggtaattgg 11700
gataacagcg agtatggacg ccagctaaaa aaatattatt caaactatcc aaatatcacc 11760

ctactagaac ctaactataa tattgaagag ctttataaaac taagaaaaaa ttgtcttgca 11820
tacattcatg gacactcgcc tggtggaaca aacccttctt tagttgaagc gatgcatttt 11880
aatattccta ttttgctt cgattgtgac tttaatcggtt acacaactaa caatttagct 11940
cattacttta atgattctga acaacttagc ttatttagcag aaagtttgc ttttgaaat 12000
cttaaatgtc gagtattaga tttaaaaaat tatgctgaag atatgtataa ctggaggcat 12060
atagctgcta tgtatgaatc tatttattaa acgcattaac aataatataa ttgacccatt 12120
atagcaggga aagatcacgt aacgctgcgg cgccgcgatc cccatatgaa tattccctt 12180
agttcctatt ccgaagttcc tattcttct agagaatagg aacttcggaa taggaactaa 12240
ggaggatatt catatggata aagccgtaag catataagca tggataagct atttatactt 12300
taataagtac tttgtatact tatttgcgaa cattccaggc cgccgacatt cagcgcggc 12360
atcacacctg acaggaggtat gtaatgtcca agcaacagat cggcgtagtc ggtatggcag 12420
tgatgggacg caaccttgcg ctcaacatcg aaagccgtgg ttataccgtc tctattttca 12480
accgttcccg tgagaagacg gaagaagtga ttgccaaaaa tccaggcaag aaactggttc 12540
cttactatac ggtgaaagag tttgtcaat ctctggaaac gcctcgtcgc atcctgttaa 12600
tggtgaaagc aggtgcaggc acggatgctg ctattgattc cctcaaacca tatctcgata 12660
aaggagacat catcattgat ggtggtaaca cttcttcca ggacactatt cgtcgtatc 12720
gtgagcttc agcagagggc tttaacttca tcggtaccgg tggcttgc ggtgaagagg 12780
ggcgctgaa aggtccttct attatgcctg gtggccagaa agaagcctat gaattggtag 12840
caccgatcct gacaaaatc gccgcgtag ctgaagacgg tgaaccatgc gttacctata 12900
ttggtgccga tggcgcaggt cactatgtga agatggtca caacggtatt gaatacggcg 12960
atatgcagct gattgctgaa gcctattctc tgcttaaagg tggcctgaac ctcaccaacg 13020
aagaactggc gcagacctt accgagtggaa ataacggtga actgaggcagt tacctgatcg 13080
acatcaccaa agatatctt accaaaaaaat atgaagacgg taactacctg gttgatgtga 13140
tcctggatga agcggctaac aaaggtaccg gtaaatggac cagccagagc gcgctggatc 13200
tcggcgaacc gctgtcgctg attaccgagt ctgtgtttgc acgttatatc tcttctctga 13260

aagatcagcg tggccgca tctaaagtgc tctctggtcc gcaaggcacag ccagcaggcg	13320
acaaggctga gttcatcgaa aaagttcgtc gtgcgtgta tctgggcaaa atcgttctt	13380
acgcccaggg cttctctcag ctgcgtgctg cgtctgaaga gtacaactgg gatctgaact	13440
acggcgaaat cgcaagatt ttccgtgctg gctgcattat ccgtgcgcag ttcctgcaga	13500
aaatcaccga tgcttatgcc gaaaatccac agatcgctaa cctgttgctg gctccgtact	13560
tcaagcaaat tgccgatgac taccaggcagg cgctgcgtga tgtcggttgc tatgcagtac	13620
agaacggtat tccggttccg accttctccg cagcggttgc ctattacgac agtaccgtg	13680
ctgctgttct gcctgcgaac ctgatccagg cacagcgtga ctatgggttgcg catactt	13740
ataagcgtat cgataaagaa ggtgtgttcc ataccgaatg gctggattaa	13790

<210> 13
<211> 13777
<212> DNA
<213> Artificial Sequence

<220>
<223> example 06A rfb locus nucleotide sequence - 06A-EPA production
strain stGVXN4112 and stLMTB10923

atgacgaatt taaaaggcgt tattcctgtt gccccgtctcg ggatgcataat gttgcctgcc	60
actaaggcga tacccaaaga gatgctacca atcgatcgaca agccaatgtat tcagttacatt	120
gttgcgttgc ttgtggctgc agggatcaaa gaaatcctcc tggtaactca cgcgtccaag	180
aacgcggctcg aaaaccacctt cgacacctct tatgagtttag aatcactcct tgagcagcgc	240
gtgaagcgtc aactgctggc ggaagtacag tccatttgcc cggccggcgt gacaattatg	300
aacgtgcgtc agggcgaacc ttttaggtttg ggccactcca ttttatgtgc acgacctgcc	360
attggtgaca atccatttgt cgtggtgctg ccagacgttg tgatcgacga cgccagcgc	420
gaccgcgtgc gctacaacct tgctgccatg attgcgcgtc tcaacgaaac gggccgcagc	480
caggtgctgg caaaacgtat gccgggtgac ctctctgaat actctgtcat ccagaccaaa	540
gagccgctgg accgcgaagg taaagtgc cgcattgttg aattcatcga aaaaccggat	600

cagccgcaga cgctggactc agacatcatg gccgttggtc gctatgtgct ttctgccat 660
atttggccgg aacttgaacg cactcagcct ggtgcatggg ggcgtattca gctgactgat 720
gccattgccg aactggcgaa aaaacagtcc gttgatgcca tgctgatgac cggcacagc 780
tacgactgcg gtaaaaaaat gggttatatg caagcgttcg tgaagtatgg actacgcaac 840
ctcaaagaag gggcgaagtt ccgtaaaggg attgagaagc tgttaagcga ataataaaa 900
tctgaccgga tgtaacggtt gataagaaaa ttataacggc agtgaagatt agcggcgaaa 960
gtaatttgtt gcgaattttc ctgccgttgt tttatataaa caatcagaat aacaacgact 1020
tagcaatagg atttcgtca aagtttcca ggatttcct tgttccaga gcggatttgt 1080
aagacaatta gcattgaat ttacgggtt tagcgcgagt gggtaacgct cgtcacatcg 1140
tagacatgca tgcagtgctc tggtagctgt aaagccaggg gcggtagcgt gctgaaatta 1200
taaagtcatcatt cttatagaac atcgcatttc aataatataa ttacacctaa atgaatagga 1260
tacaacgtgt gcacaattat ttaaggctta aagataaaat aaaaaacgta tttttagggt 1320
tgtatatattt gcagtttattt aattatatcg cgccatttgtt aattatccct atcctgataa 1380
aatatattgg gttggggaa tatgggaat tggctatat tacatctatt tatcaaata 1440
tggcttgat tattgatttt ggcttactt acacaggacc tgtggttgct gcgagacata 1500
gatgtgagac ccaaaattta cagcgctatt actcaatagt tggctttta aaatcattgc 1560
tttttataat tgcattaaca tgtgtatttt tattgtgcag attaaatata gtccacttgt 1620
catttttgg gttttgtca attttctat gcactattgg taatatatta tcgcccatt 1680
ggttttgca ggggatttgtt gatTTaaaa aactttcata ctcacaagta atagtgagaa 1740
taacattgtt tatcatactt cttgtttatg tctgttagtgg cggagataat gtttttatcc 1800
taagttttt gcaaaatgca acattactca tatgctgtat atacttatgg ccaaataattc 1860
atattagcca tggttttcat cttaaaccta atgaatgcat tgtggaaattt aagaaggcag 1920
gaaatgtttt tattggcgta ataggtacga ttggttacaa tggcttaatt cctgtgttaa 1980
ttggaaacct ttgcggtaat acgagtcttg gtgtttttc aatcgaaat aaaaatgacaa 2040
cagcatgtca aagtctaatt aatccaaat cacagtataat gttatctcaa gtttcagaaa 2100

ttaaacctca agataaactg ttttattata gaattaaaaa aagtttttt gtgcatttaa	2160
caattagcat aattgcatgt ttatgttata tggggttagg gcaatatgtg gcgactttta	2220
tagttaagt tgacgttca tttgttatta ttttatttgc gtcaataatt accattttt	2280
catcttaaa taatgtcctt ggtatacagt ttcttataacc gacagataat gtaaaaatac	2340
tacgaagtat aaatgttatg gcgggaatta ttgtttagt tttgtcctgg ctgttaatat	2400
cacgcttga cattctgggg ggggtttat taaacctaattt tggtgagttt cttgtattca	2460
gtatgcttagc ttttattgcc catcgaaagt ggggaggcag agtataatga aagtgaaggc	2520
ggttcctgct attacattct atttaagttt aatgctgaca atttttagtgt tactgtttgg	2580
taatgaacca aataaatcac aatatacct ttttatagca acgataacag ttttttat	2640
cgcataatc actaataaaaa taacttctcc ggccagccctt ctcgttatcatctttgt	2700
gttttaggt tgtcgccctt tattatcttt gtttgcaaac tatgattata ggattgccga	2760
ttggtttattt gaaggatata tggatgacga tgtgattttg gctaactatg ctataacact	2820
aatgtattat ggttatacat tggactaat tctatgaaa aatactgaaa aattttatcc	2880
gcatggcct tattctgaaa aacaattgct aaaaataaaag tttctttga ctttat	2940
tctgggttcg ataggtatgg ttgtaaaagg gatattcttt tttaacttta tagaatctaa	3000
tagttatgtt gatatttatac aatcaaataat aacaacgcca ataggtttagt attttctatc	3060
ttattttattt tattgttctt tttccttat atgtgcgtt catatacgt tcagaacaaa	3120
taaaaaattt cttttatttgcgatatgcat tgctgcattt agcacccgttga aggtagtcg	3180
tagtgaagct ataacgttcc ttttaacggt tacatgtata tattttatg aagtaaagac	3240
aagaaaactta cgtctgctga ttacaatgat ttttgtttt agcgtcattt ttgtgattag	3300
tgaattttatc tcaatgtggc gcactggagg gagttttttt caattaatgc agggtaataa	3360
tcctgttata aactttgtat acggcatggg agtacatcatat cttccattt atcaatcgt	3420
aaaactacaa ctattgtcag gggatataa tgttacctat ctattcagcc agttaataat	3480
aacttgctcg tcaatattta atgtcaaattt gagcttgccg gaaataagct atagccattt	3540
ggcctcatac acagcaaacc cagaactata taatcttggg ttggacttg gggggagttt	3600

ttagcagaa tcgttttag cattggct gattggatgt ttcattatac ccttttact	3660
tttacttaat taaatgtat tgaaaaata tacaaaaaac aaaccaatta tatattttgt	3720
ttattatagt gtgtgccac ctatattatt cacaccaaga gagacttgt tctatttctt	3780
cccttatctt gtcaaagta tattgtgc ttttttagtt acattataca tccagtataa	3840
aaaggattga ccaaattgtc agaaaaaat gtcagcataa taatcccaag ttataacagg	3900
gctcatattc ttaaggaggt cataccaagt tatttcagg atgagacttt agaggtata	3960
gttatcaatg atggatcaac agataataca aatagtgtat tagctgaact gaaggaaaaa	4020
tatttcagt tagttttt agaaaatgaa acgaataaaa aacagatgta ttctaaaaac	4080
cggggattg aaatagccaa agggaaatat atttttttg gtgatgatga ctcttacctc	4140
ttacccggtg ttatatctcg gttattggct acaaaatatg agacaggcgc tcatgtatc	4200
ggcgaagaa tactttat gaataataac gagaaaacaa ttgaagattg cataaattcga	4260
cataaaaaag agggcgttt tgtagtgat ctaaatagat tggatttttag ttatacatgt	4320
gatttggacc atccgattga atgttttat gcacagccctt ttgttcttagc tggaaaggaa	4380
ctaataatcga aatatcgatt tgatatatct tatacggaa actgctatcg tgagggaaact	4440
gatttcatgc tatctctatt tattaaaat aaaaattta tatacgatttca aaggctttg	4500
ttaataaatt tacctccaag aaaagcgacg ggaggggcaa gaacagctaa tcgattaaaa	4560
tatcattacg aaagttgcatt aaataattat agattttaa aaaaatataa tgataatttg	4620
aatctcttt caggacaaaa gcatgctata ttttaccgac agtgcattt cggtctgcta	4680
aaaatgaagt cgtttatcgg gaagttttta aaatgattat atatatcgcc gcgtataatg	4740
gttcaggagg gcaagggtggg gtggaaaggg ttgttgccta acaatgtaac attctaaaa	4800
atttgggggt taaagtcatt atacttgata aaacatactt caaaatttct aacaaaattc	4860
gtaacaaaaa aatacaagta gcactttatc caatattgtt ttctctttat ttaaccttac	4920
aaaaattacg tggcgtgacg tttaaagtta ttgcacatgg ctattgttct ccttttata	4980
ggaatgacat cttaatagct catggcaata tgaaatgtta ttttcaaaca gtcgtatgt	5040
aaaaacctaa tcggttgtct ggcagtggc ttttatctt ctatgagcgt tggcgtggag	5100

cattttcaaa aaatatctgg gctgttcaa ataaggtaa aagtgaatgg aatgagctt	5160
acaatattaa ttcacataaa atcaaagttg ttcgaaattt tataaatctt gcacaatttg	5220
attacactga tgttaatgaa gcagaatatg tgacatttgt cggcgattg gaaaaaggaa	5280
aaggaataga tgatctgtat tacatatgta aaaatctgcc agataacttcc ttccattnag	5340
tttcaagtat tcccgcucca caaaattttg ctgcgtaaa taatgttctg accagcattg	5400
ctgtcccccta tgcgaaaatg ccagaaatat ttaagaaatc cagagtactt attttaccgt	5460
cctattatga aggatatgag ctggttacta ttgaaggcgt atgctgtggt tgccctgtga	5520
taggctataa tggtggtgca attagagagt tgtatgcaga aagtttcct ggcgtattta	5580
ttgccaataa taaagaagat ttagcacaag tagcctacaa attaatttagt cttgataatg	5640
aaaaatatta tcatttgaga caaactattt atagcaagcg tgagttttt tctgaagaga	5700
gatatgcgga aattttaacg gcggcattta atgaaaaaaaaa ataagaaact ctgtctcatt	5760
tcaattaact catataatga acttaccgga ggaggaggtt attacgtac gcttggtagt	5820
tttctacaaa aacagaatgt taatttaaca cttattgata aaaaatcctc aggtaaacta	5880
ttcgaagaca atactttca acatatatca tttattaaag gtaaacgtca ggatataata	5940
tccaggcttt ttttataacc atcattttat gtcccttata ttttctcaat aattaaaatt	6000
ttacggaagc aagatattct tgctttcac aactctcggc ttggattgtt atgtctgctt	6060
tttagaatac tcatgcucca caaaaagatc atattgtta cgataactt cgaatatgac	6120
ttaataagac aaaaagataa aaacataact actttattg aaaaattaat tgtttatctc	6180
aatgaattta tcgggcttaa gaattcagat ttagttagct atattacccg gcaagataaa	6240
aatgcaatgg ataaattttt tgggattaaa aaaagcagaa attaattct ccctgtgata	6300
tttagtagag aaaaaccaac tgatgtattg tcagctcact ttattaaatga gtataatcga	6360
ttgaataatg ataataggaa aaaagtagta ttactgcat ctttgattt ttttccaaat	6420
atagatgctg ccaactatgt tttaaatgca gcaaagtcta ataatgatta ttgctatatt	6480
ttggcaggtt gaaaaagtac tacttgaat cttcctgatt tgataattt atttttttc	6540
gataatctat ctaatagtga aatgtcatat ttattatctg cttgtatgt ttttattct	6600

cctatagttt taggaagtgg aatgaaaaca aaaattgcag aagcactatc atatggatta 6660
tatattttatg cgacagagca ttccttaatc ggctatgatg aaattataca caataaggag 6720
tgtgttaaaa aaatctcaca tttggatgag gaatttccta aagatttcaa gatgaaaagt 6780
atcaataaac agctaataat gtcttatcag caaaaatattt attcacatta tcggttaat 6840
ggccatgaac ttgatataat aaatttgac gattagttag tggagatata atatgaacat 6900
attagtaact ggtggtgctg gatatatcg atctcatacg gctattgaat tactgaatgc 6960
aggtcatgag attatcggttc tggacaattt cagtaatgct tcatacaagt gtatcgaaaa 7020
aataaaagaa attactcgac gtgattttat aacaattact ggagatgctg ggtgttagaa 7080
gacactctcc gctatttcg agaaacacgc catagatata gttattcatt ttgctggctt 7140
taaatctgtt tcagagtcta aaagtgaacc cttaaagtat taccagaata atgttggagt 7200
gaccattact ttattacagg taatggaaga gtacagaatt aaaaaattta tcttagtttc 7260
atctgcgaca gtctatggtg aaccagagat aattccaatt ccagaaacag ctaaaattgg 7320
aggaactacg aatccatatg gcacatcgaa gtattttgtt gaaaaatttc tagaggatgt 7380
tagttccacg gaaaaactgg atataatttg cttgagatat tttaatcctg tcggtgctca 7440
ttcttagtggt aaaataggtg aggctccatc tggtatccct aataatcttgc ttccattttt 7500
attggatgtt gcgagtggtt aacgtgataa attattttatt tatggcaatg attaccctac 7560
taatgatgga acaggtgtaa gggattttat tcatgttgc gacttagcga aaggtcattt 7620
ggctgcaatg aattatttaa gtatcaattc gggatataat atcttaatc ttggtagagg 7680
aaaaggttat tcggtacttg aattaatcac tacatttgc aaattaacaa acattaaggt 7740
caataaatct tttatagaga gaagggcagg ggatgttgcg tcttgggg ctgatgcaga 7800
taaagctaatttcttattgg actggcaagc cgaacaaact ctagaacaga tggttattgg 7860
ctcgtggcgt tgaaaaaaa attatccaga cggattctga atataaaagg ttgcgttttt 7920
atgaatcaat cagagcagag aaaaaaaaaa ctggttctt cacctcgctt tccctaccct 7980
gtcattggag gggatagatt aagagtctat atgttgcgtt aagaactttc caaaaatata 8040
gatcttatttgc ttctgagttt atgtgatcaa ccactagaac ttgaaataaa tataaatgac 8100

tcggcttca aagaaattca tcgtgtctat ctacaaaaat ataaatcata ttataatgta	8160
ttaaaagctt tggttacgca aaaaccgttg caaattgctt attatcaatc ggacacattt	8220
aagaataaat acaataaatt aattaaacaa tgcgatgcag tatttgtca tctgataaga	8280
gttgctgatt atgttaagga tacagacaag ttcaaaattc ttgatatgac agatgcaata	8340
tcttgaatt acagtcgcgt taaaaaatta gcaagtaaaa aaagtttgcg tgcaattatt	8400
tattctctgg aacaaaaaaag attagaatca tatgaacgtt ctgtggcgaa tcttttgat	8460
ttgaccactt ttatttcatc cgtagaccgt gactatctc accctaattct gggcagtaat	8520
atccatatacg tcaataatgg gggtgataca tcagccttga gatatataaa aagagaaata	8580
aaaatcgata agcctgtgga acttatattt atcggaaata tgtattctt acaaataatg	8640
gatgctgcaa aacattttgc taagaatatt ttaccttgct tgtatgatga gtttaatatt	8700
attttaaag tgattggtaa gatctcagaa actaataaaa atatattaaa ttcatttaaa	8760
aatacaatttgc ctttaggtac tggtgatgat atcaattctt ccgcttctac agggcatata	8820
ggtatatgtc ctgttcgtct tggaggcaggc gtacaaaata aaattcttga atacatggct	8880
ttagtttac catgtattac atctagcatt ggttatgaag gtattatgc aaaatcaggt	8940
agcggaaattt ttgttgcaga tacagtagag caatataaaa acgtactaag agaaataatt	9000
tacgattata atcgttatac tgaagtggct gaaaatgccc gtatggctt agaaaataat	9060
ttttcttggg aatcaaaagt tgccaatttta atgaatacat tagatgagaa attatatgaa	9120
caataataaa attattacac ctatcattat ggctgggtt tcaggcagtc ggttggcc	9180
actatcaaga attctctatc cgaaacaatt tcttagccta atcggtagtc ataccatgct	9240
tcaaacaacg gctaattcgtc tggatggttt ggattgtacc aaccctttag tcatttgtaa	9300
tgaacaatac cgctttatag ttgctgaaca gcttagaaaa atcgatagat tgacttcaaa	9360
gaatatcatc cttgagcctg ttggcgtaa cactgcccct gcaattgcat tagcggcggtt	9420
gctgatgtct aagtctgata aaagtgcaga tgatctttagt ctcgtactgg ctgcagatca	9480
cgttatacac gatgaagaaa aattttgtaa cgctgtttaga tcggcaattc catacgctgc	9540
tgtatggaaaa ttggtaacat ttggtataat tccagacaaa gcagaaactg gttatggta	9600

tatacatcga ggacaatata ttaatcagga agattcgat gcatttatag tgtcatcatt 9660
tgttggaaag ccaaatacatg agacagccac taaatatctt gcttccggtg agtattattg 9720
aatagcggt atgttttgt ttagtgcaaa tcgttatata gaggaactta aacaatttcg 9780
gcctgatatt ttatccgctt gtgaaaaagc aattgcttca gctaactttg accttgattt 9840
tgtgcgttta gatgaaagtt ctctctctaa gtgccctgaa gaatcaattt attacgctgt 9900
aatggaaaaa acaaaagacg caattgttat tccaatggat gctggctgga gtgatgtcgg 9960
ttcatggtct tctctttggg aaattaatga taaagactca gacggcaacg taatagttgg 10020
ggatatttc tctcatgaaa caaagaattc tttcatatat gccgaatcgg gaattgttgc 10080
tacagttgga gtggaaaatt tagttgtgt ccaaacaag gatgctgttc ttgtctcaga 10140
gagaaataaa gttcaggatg taaagaaaat agtagaacaa attaaaaatt caggtcgtag 10200
cgagcattat gttcatcgcg aagtatatcg tccttgggt aaatatgatt ccattgacac 10260
aggggagcgt tatcaggtca aacgtataac agtaaatcct ggtgaaggac tttcttaca 10320
aatgcaccat catagggcag aacattggat catagtttct ggaactgcaa ggggtgactat 10380
aggttctgaa actaagattc ttagcgaaaa tgaatctgtt tacatacctc ttgggtgtaat 10440
acactgcttg gaaaatccag ggaaaattcc tcttgattta attgaagttc gttctggatc 10500
ttatttagaa gaagacgatg ttatccgttt tcaggaccga tatggtcgta gctaaatttt 10560
tgataatgtt acgttagtag aagagcgcta atatttttag ttaatctgtt ataagtatta 10620
tttggtaag gtatcatg tcgagttac cctgctttaa agcctatgtat attcggggaa 10680
aattaggcga agaactgaat gaagatattt cctggcgcat tggtcgcgt tatggcgaat 10740
ttctcaaacc gaaaaccatt gtgttaggcg gtgacgtccg actcaccagc gaaaccttaa 10800
aactggcgct ggcgaagggg ttacaggatg cgggcgtcga tgtgctggat attggcatgt 10860
ccggcaccga agagatctat ttgccacgt tccatctcgg cgtggatggc ggcacatcgaag 10920
ttaccgcccag ccataacccg atggattaca acggcatgaa actgggtgcgc gaaggggctc 10980
gccccatcag cggtgataacc ggactgcgcg acatccagcg tctggcagaa gccaacgact 11040
ttcctccgt tgatgaaacc aaacgcggtc gctatcagca aatcaatctg cgtgacgctt 11100

acgttgcata cctgttcgg tatacaacg tcaaaaacct cacggcgctc aagctggta 11160
ttaactccgg gaacggcgcg gcgggtccgg tggtgacgc cattgaagcc cgctttaag 11220
ccctcggcgc acccgtaaa ttaatcaaag tgacacaacac gccggacggc aattttcca 11280
acggtattcc taacccgcta ctgccgaat gtcgcacga caccgcata gcggtcatca 11340
aacacggcgc ggatatggc attgccttg atggcgattt tgaccgctgt ttcctgtttg 11400
acgaaaaagg gcagtttatt gagggctact acattgtcgg cctgctggca gaagcggtcc 11460
tcgaaaaaaaaa tcccggcgc aagatcatcc acgatccacg tctctcctgg aacaccgtt 11520
atgtggtgac tgccgcaggc ggcacccgg taatgtcga aaccggacac gccttattt 11580
aagaacgtat ggcgaaggaa gacgctatct acggtggcga aatgagcgcc caccattact 11640
tccgtgattt cgcttactgc gacagcggca tgatccgtg gctgctggc gccgaactgg 11700
tgtgcctgaa aggaaaaacg ctggcgaac tggtgccgca ccggatggca gcgtttccgg 11760
caagcggtga gatcaacagc aaactggcac accccgttga ggcgattaac cgcgtgaaac 11820
agcactttag ccgcgaggcg ctggcggtgg atcgacccga tggcatcagc atgacccttg 11880
ccgactggcg cttaacctg cgctcctcta acaccgaacc ggtggtgcgg ttgaatgtgg 11940
aatcgcgccg cgatgtaccg ctgatggaa aaaagacaaa acttacccctt gagttactga 12000
acaagtaatt cagtaatttc atataatgg gttttaaaaa acggaaaaga tgagatattcc 12060
ggtgtggtat atccaaggta atgctattca gtatcttat gagtgagtta acatctatac 12120
cacatthaag ccgcacactt cgggatcccc atatgaatat cctccttagt tcctattccg 12180
aagttcctat tctttctaga gaataggaac ttccgaatag gaactaagga ggatattcat 12240
atggataaaag ccgtaagcat ataagcatgg ataagctatt tatactttaa taagtacttt 12300
gtatacttat ttgcgaacat tccaggccgc gaggattcag cgggtgatc acacctgaca 12360
ggagtagtta atgtccaaagc aacagatcg cgtagtcgg atggcagtga tgggacgcaa 12420
ccttcgcgtc aacatcgaaa gccgtggta taccgtctt atttcaacc gttccgtga 12480
gaagacggaa gaagtgattt ccgaaaatcc aggcaagaaa ctggttcctt actatacggt 12540
gaaagagttt gtcgaatctc tggaaacgcc tcgtcgcatc ctgttaatgg tgaaagcagg 12600

tgcaggcacg gatgctgcta ttgattccct caaaccatat ctcgataaaag gagacatcat	12660
cattgatggt ggtaaacacct tcttccagga cactattcgt cgtaatcgtg agctttcagc	12720
agagggctt aacttcatcg gtaccggtgt ttctggcggt gaagaggggg cgctgaaagg	12780
tccttctatt atgcctggtg gccagaaaaga agcctatgaa ttggtagcac cgatcctgac	12840
caaaatcgcc gccgtagctg aagacggtga accatgcgtt acctatattg gtgccgatgg	12900
cgcaggtcac tatgtgaaga tggttcacaa cggtattgaa tacggcgata tgcagctgat	12960
tgctgaagcc tattctctgc ttaaagggtgg cctgaacctc accaacgaag aactggcgca	13020
gaccttacc gagtggaaata acggtaact gagcagttac ctgatcgaca tcaccaaaga	13080
tatcttcacc aaaaaagatg aagacggtaa ctacctggtt gatgtgatcc tggatgaagc	13140
ggctaacaaa ggtaccggta aatggaccag ccagagcgcg ctggatctcg gcgaaccgct	13200
gtcgctgatt accgagtctg tgtttgcacg ttatatctct tctctgaaag atcagcgtgt	13260
tgccgcatct aaagttctct ctggtccgca agcacagcca gcaggcgaca aggctgagtt	13320
catcgaaaaa gttcgtcgtg cgctgtatct gggcaaaatc gtttcttacg cccagggctt	13380
ctctcagctg cgtgctgcgt ctgaagagta caactgggat ctgaactacg gcgaaatcgc	13440
gaagattttc cgtgctggct gcatcatccg tgcgcagttc ctgcagaaaa tcaccgatgc	13500
ttatgccaa aatccacaga tcgctaacct gttgctggct ccgtacttca agcaaattgc	13560
cgtactac cagcagggcgc tgcgtgatgt cggtgcttat gcagtacaga acggtattcc	13620
ggttccgacc ttctccgcag cggttgccta ttacgacagc taccgtgctg ctgttctgcc	13680
tgcgaacctg atccaggcac agcgtgacta ttttggtgcg catacttata agcgtatcga	13740
taaagaaggt gtgttccata ccgaatggct ggattaa	13777

<210> 14
 <211> 15027
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> example 08 rfb locus nucleotide sequence - 08-EPA production
 strain stLMTB11734

<400> 14
atgacgaatt taaaaggcagt tattcctgtta gcgggtctcg ggatgcataat gttgcctgcc 60
actaaggcga tacccaaaga gatgctacca atcgtcgaca agccaatgtat tcagtagtatt
gttgacgaga ttgtggctgc agggatcaaa gaaatcctcc tgtaactca cgcgtccaag 120
aacgcggctcg aaaaccacctt cgacacctct tatgagtttag aatcactcct tgagcagcgc
gtgaagcgtc aactgctggc ggaagtacag tccatctgtc cggcgggcgt gaccattatg 240
aacgtgcgtc agggcgaacc ttttagttta ggccactcca ttttgtgtgc ggcacccgtcc 300
attggtgaca acccatttgt cgtggtaactg ccagacgttgc tgatcgacga tgccagcgc
gaccgcgtac gttacaacct tgctgcatg attgcacgtt tcaacgaaac gggccgcagc 420
caggtgctgg caaaacgtat gccgggtgac ctctctgaat actccgtcat ccagactaaa 480
gagccgctgg accgtgaggg taaagtcaacg cgcattgttgc aatttatcga aaaaccggat 540
cagccgcaga cgctggactc agacatcatg gccgttaggtc gctatgtgct ttctggcgat 600
atttggccgg aactggaacg tactcagcct ggtgcattgg gacgtattca gctgactgtat 660
gctattgccc agctggcgaa aaaacaatcc gttgatgcaa tgctgatgac cggcgcacagt 720
tagcactgca gcaaaaaat gggctatatg caggcgttttgc tgaagtatgg cctacgcaac 780
ctgaaagaag gggcgaagtt ccgtaaaggt attgagaagc tgtaagcga ataatgaaaa 840
tctgaccgga tgtaacggtt gataagaaaa ttataacggc agtggaaattt cgcagcaaaa 900
gtatattgtt gcgaatcttc ctggcgttgtt tttatataaa ccatcagaat aacaacgagt 960
tagcagtagg gttttattca aagtttcca ggatttccct tggttccaga gcggattgg 1020
aagacaatta gcgttgaat tttcgggtt tagcgcgagt gggtaacgct cgtcacatca 1080
taggcattgc tgcagtgc tggtagctgt aaagccaggg gcggtagcgt gcattaaatc 1140
ctctattaaat caaaactgaga gccgcttatt tcacaggatg ctctgaagta atatgaaata 1200
aattaagcta gcgatcgctt aagatctagg atttcattat gttacttcct gtaattatgg 1260
ctgggtgtac cggcagtcgt ctctggccga tgtcacgcga gctttatccg aaacagttcc 1320
tccgcctgtt cggcagaac tccatgctgc agggaaaccat caccgcactc tcggcccttg 1380
aaatccatga accgatggtc atctgttaacg aagagcaccg cttcctgggtg gctgaacagc 1440
1500

tacgccagct caataagctg tcgaataata ttattcttga gccggtcggg cgcaacaccg 1560
ccccggccat cgccctggca gcccttcagg ccacccgcga cggcgacgac ccgctgatgc 1620
tggttctcgc cgctgaccat atcatcaata accagtcggc cttccacgac gccatccggg 1680
tcgcccggca gtatgctgat gaaggtcatc tggtcacctt cggtatcgtg ccgaatgccc 1740
cgaaaaactgg ctacggttac attcagcgcg gcgtggcgct caccgatagt gcccattccg 1800
cgtaccaggt ggcccgctt gtggagaagc cggatgcga gcgcgcccggag gcttacctcg 1860
cctccgggga gtactactgg aacagcggca tggttatgtt ccgcgccaag aaataaccta 1920
tcgagctggc caaataccgt ccggatatcc tggaaagcctg ccaggctgcg gtgaatgccc 1980
ccgataatgg cagcgatttc atcaatatcc cgcatgatat tttctgcgag tgcccgatg 2040
agtccgtgga ctatgccgtt atggagaaaa ccgcccgtgc ggtggtggtc ggtctcgatg 2100
ctgactggag cgacgtcgcc tcctggtccg cactatggga ggtcagcccg aaagacgagc 2160
agggcaatgt cctcagcggt gacgcgtggg tacacaacag cggaaaactgc tacatcaaca 2220
gcgacgagaa gctagtggcg gccattggcg tagagaatct ggtgattgtc agcactaagg 2280
acgcccgtgct ggtgatgaat cgcgagcggtt cccaggacgt gaagaaggcg gtcgagttcc 2340
tcaagcagaa ccagcgcagc gagtacaagc gccaccgtga gatttaccgc ccctggggcc 2400
gttgcgacgt agtggtccag accccgcgt tcaacgtcaa ccgcacacg gtgaaaccag 2460
gcgggtgcctt ctcgatgcag atgcaccacc atcgcgcccga gcattgggtt attctcgccg 2520
gcaccggta ggtgactgtc aacggtaagc agttccgtt gtccgagaac cagtccacct 2580
ttattccgat tggcgccgag cactgcctgg aaaaccctgg ctgtattccg ctggaaagtgc 2640
tggagatcca gtcggggcg taccttggcg aggacgacat tattcgtatt aaagaccagt 2700
atggtcgttgc ctaattatcc tcgggacaag acgcagaatg acacagttaa cttgtttaa 2760
agcttatgac atccgtggtg aactgggtga ggaactgaac gaggacatcg cctaccgtat 2820
cggtcgcgcc tacggcgaat ttctgaaacc cgggaagata gtgggtgggg gcgatgtgcg 2880
cctcacaagc gagtcgctga agctggcgct ggcccgccgg ttaatggacg ccggtaccga 2940
cgtgctggac atcggcctga gcggtaccga agagattac tttgccaccc tccacccgttgg 3000

ggtagatggt ggcatcgagg tgaccgcgag ccacaatcct atgaactaca acggcatgaa 3060
gctggtgcbc gagaatgcga agcccatcag cggcgacacc ggctgcggg atatccagcg 3120
cctggcggag gaaaaccagt tcccgccagt ggacccggcg cgtgcggga ccctgagcaa 3180
gatatcggtt ctgaaggagt atgttgcacca tctgatgagc tacgtggact tctcaactt 3240
caccgtcca ctgaagttgg tggtaactc cgaaacggg gctgcggggc acgtgattga 3300
tgaggtggag aaacgcttcg cggcgctgg ggtgccggta accttatca aggtgcatca 3360
ccagccggat ggccatttcc ctaacggtat cccgaatccg ctgctgccgg agtgcgcaca 3420
ggataccgcc gacgcggtgc gcgagcatca ggccgacatg gggattgcct ttgacggcga 3480
cttcgatcgc tgcttcgtt tcgatgacga agcttcgtt atcgaggggt attacattgt 3540
cggcctgctg gctgaggcgt tcctgcagaa gcagccggg gcaaaaatca ttcacgaccc 3600
gcgcggatc tggaacacgg tagacatcgt gacccgcaac ggcggccagc cggtgatgtc 3660
gaagacgggg catgcgttca tcaaggagcg gatgcgtcag gaagacgcta tctacggcgg 3720
ggagatgagt gcgcaccatt acttccgcga tttcgctac tgcgatagcg ggatgatccc 3780
gtggctgctg gtggcgagc tgctgtgtc gaagaacagc tcgctgaaat cgctggggc 3840
ggaccgcag aaggcggttcc ctgcgtcggg agagatcaac cgcaagctaa gtaatgctgc 3900
tgaggcgatc gcccgcattcc gggcgagta tgagccggcg gctgcacaca tcgacacaac 3960
ggacgggatc agtattgaat accctgaatg gcgcggatc ctgcgcacgt ctaacaccga 4020
gcgcggatc cgtctgaacg ttgagtccag agctgatgtc ggccttatga ataaaaaaac 4080
gaccgagctg ttacacctgt taagcggggtaa agatgtgag agatttacta acgacgattt 4140
atcggttatcg gggatttatac tggagcagtg ttaaacgtga tttcaggca cgctatcaa 4200
ctagtatgct gggcgacta tggctcggtt tacaaccgct ctctatgatt ctggctata 4260
ccctgggttt ttccgagggtg atgaaggcaa gaatgcccga taataccggg tcgtttgcct 4320
atagtattta tctctgttcc ggggtactga cctggggatt atttactgag atgctggata 4380
aaggtcagag cgtatattt aacaatgcta atctgatcaa gaaactcgtt tttccaaaa 4440
tctgtctgcc gatcatcgatc acgttatcgg cggtgcataaa tttcgcgatt atttcagtc 4500

tgtttctaat ttttatcatt gtcaccggta acttccccgg ctggctctt ctctcggtga 4560
taccggtcct gctttgcag atcctgttg ccgggtggct ggggatgatc cttggtgtca 4620
tgaacgtctt tttcagggat gtggggcaac tgggtggcgt tgccgtgcaa ttctgggttt 4680
ggttcacacc cattgtttat gtactgaatt cattacctgc atgggcaaaa aatctgatga 4740
tgtataaccc gatgactcgg atcatgaaat cttatcagtc catctcgcc tatcatctgg 4800
cccccaactg gtattcgcta tggccagttat tggctctcgc cattattttc tgcgtcatcg 4860
gtttcaggat gttccgcaag catgcggcgg atatggtgga tgaattataa tgagttatat 4920
cagagtaaat aatgtcggta aggctatcg ccagtatcac tcaaagaccg ggagactgat 4980
cgaatggtta tcccctctga ataccaaacg ccataatttgc aaatggatcc tccgcgatata 5040
taatttcgaa gtcgctccgg gcgaggctgt cggtattatc ggtatcaacg gtgcaggcaa 5100
gagtagccctg cttaaactca taaccgggac gtccaggccg acgactggag aaattgaaat 5160
ctccggacgt gtcgctgcat tactcgaatt ggggatgggg tttcattctg atttcactgg 5220
tcggcagaat gtttatatgt ctggcaact gttgggtta tcgtcagaga aaataactga 5280
actgatgccg caaattgaag agtttgcata gattggggac tatatcgatc aacctgtgcg 5340
cgtctactcc agtgggatgc aagttcgatt agcttttagt gtagcgacgg ctatccgtcc 5400
tgatgtgcta attatcgatg aggattatc tggggat gcatatttcc agcataaaag 5460
ctttgagcgt attcgaaaat ttgcgtcagga agggaccacg ctgttgctgg tatccatga 5520
taaacaagcg atccaaagca ttgcgaccg ggccatttttgc ttgaataaaag gccaaattga 5580
aatggaaggt gaacctgaag cagtgtatggc ttattacaat gctttctgg ccgataaaaca 5640
aaatcagtcc attaaacaag ttgagcataa tggtaaaaacg caaactgttt caggcactgg 5700
tgaggtgact atctctgagg ttcatcttct cgtgaacag ggcaatgtga ctgaattttgt 5760
ttcggttaggg catcggtcga gcttgcaggt caacgttgag gtcaaggacg atattcctga 5820
gcttgggtc ggatatatgat ttaaggatcg acttgggcag ccgattttcg ggaccaatac 5880
gtaccatctc aatcagacac tcacccctt gaaaaaagga gaaaagcggtt cggttcttatt 5940
ttcttcgat gcgagattgg gggttggctc ctattctgtc gctgtcgcgt tgcataacttc 6000

cagtacgcac ctcggcaaaa actatgaatg gcgcgatctg gccgtggtat tcaacgtcgt 6060
taaacggaa caacaagagt ttgtcggcgt gtcctggttg ccgcctgaac tggagatttc 6120
ttaatgggtt cgtcgtttta tcgttcattt gaagaacgac acagaggttc gggtgaagaa 6180
atcaagcgcc gcttgagttt ttattnacct tttcttgcag gtctgaagga catttacct 6240
gatggcgtga ttgcggatat tgggtgcgga cgtggcgaat gggtggagat cctgactgaa 6300
aatggcattg cgaacatcgg cgtcgatctc gatgatggca tgctggcgcg cgccagggag 6360
gccggactga atgtgcagaa aatggattgt ctgcagttt tgcaaagtca ggcggatcag 6420
agcctgatag cggtgaccgg tttcatatt gctgaggcatt tgccgtttga ggccctgcag 6480
caactcgcca tgcataccct acgggtgctg aaaccaggtg gtttgcgtat cctcgaaacg 6540
ccgaaccgg agaatgtaag cgtcggcacc tggtcatttt atatggatcc aacgcataat 6600
catcctctgc caccgcccact gcttgagttt ttacctattc attatggttt tacccgagca 6660
attaccgttc gtctgcagga aaaagaggtt cttcaatctc cgatgcagc cgttaatttgc 6720
gtcgatgtac tcaaagggtt gagccccgac tacagcatca ttgctcagaa agcagcgcca 6780
acagatattc ttgaacgctt tgacaccctg tttaccagc agtacggct gacgctggat 6840
gctctgagca accgttacga tgcgattttg cgccaaacagt tttcgtccgt tgtctcacgg 6900
ctggagacgt tgaaccaaac ctatatgcaa cagataagcc aaatgtcaga gactattcag 6960
acgttgcaag gtgaggttga cgatctgagt catgtcatcg atcagaacca tcagcttcat 7020
cagcaaatgg cggatttaca taacagtcgt tcatggcgtt ttactcaacc actacgctgg 7080
ttgtcttgc aacgtcaatt attacgtcag gaaggggcta aagtgcgagc ccgttagggct 7140
ggggaaaaaaa tattgcgcaa agggatggcg ctctcgctgg tcttttcca tcgttaccct 7200
aagtctaagg tttatctgtt taaggttctg agaaaaactg gctgctatac attgctacaa 7260
cgtttggcc aacgcgtaat gctggtgcaa tctgacacga tggatgtcgtt gtccagaaga 7320
tatgtatgtgg gtactgaaga aatgacaagt cgcgatgcgtt gtatttataa cgaattaaaa 7380
aataaaaaata cggagaaata acgatgcgtt ttgtcataga tttacaaggc gcacagacgg 7440
aaagccgctt tcgtggcattc ggtcggtata gtatgcgtt cggcagaggc ataatcagaa 7500

ataacagccg gcatgagatt ttcatgcgc tatccgccat gctggatgag tcgattgcaa 7560
atattaaggc gcaatttgcc gatctcctgc cggcagaaaa tatagtcgta tggcatgccg 7620
taggccctgt tcgtgcgatg gaccaaggta atgaatggcg tcgggagagc gcagaactga 7680
ttcgggaagc gtttcttcaa tcattgtgtc cagatgtcgt tttcattacg agtttgggg 7740
aaggcatgt cgacgatgctg gctacatcg tacacaaatt tagtcgtcag tataaagtag 7800
ccgtactgca ccacgatctt atccccctcg tgcagggga aacctatctg caggacgatg 7860
tatacaaacc ctactattta cagaaagttg agtggtaaaa aaacgctgac cttttgttga 7920
ctaactctgc ttataccgca caggaagcga tcgagcatct gcatttacag ggcgatcatg 7980
tgcagaatat tgccggca gtcgattctc agttttgtat ggcggaggtg gcagcgagcg 8040
aaaaagagac cgtccttggc cattacggta ttcagcgcga gttcatgttg tatgcgcccg 8100
gaggatttga ctcaaggaaa aactttaaac gggtgattga ggcctatgcc gggctcagtg 8160
atgccttacg tcgcagtcat caactggta tcgtcagtaa gctttccatc ggtgatcg 8220
agtatctgga atcccttgcg tcaggtaatg gtttacagca gggcgaactg gtactcactg 8280
gttatgtgcc ggaagatgag ctgatccagc tctatgcct atgtaagctg ttcatcttg 8340
cttcactaca tgaaggttt ggggtgccgg ttctggaaagc aatgtcgtgc ggtgcgccgg 8400
tgattggctc aaatgtcacc agtattcctg aagtcatcg taatcctgag gcattattcg 8460
acccgtattc tgtctttcc atgagggata agatcgcgcg atgtttgcact gatgataacct 8520
tcctcgccgc tctgaaagaa atggcgcagc agcaaggcgta taatttctct tgggataaag 8580
ctgcggtgac tgctctggaa gctttcgaaa agatcgcggt agaagacacc ggtactgcgc 8640
aggtttgcc tgaagcttg attcagaaga tccttgctat ctcacaaggg cagccagatg 8700
accgcgatct gcgcttgtgc gcaacggcca ttgattacaa tctgaaaacg gcagaacttt 8760
atcaaatcga cgataaatcg ctgaactggc gtgtggagg cccattcgat agctcatata 8820
gtctggcggtt ggtcaaccgc gaatttgcgg gggcactctc agccgatggt gtagaggttt 8880
tattgcattc cactgaagga ccaggtgatt ttgccccaga tgccctcgat atggcacagt 8940
cgaaaaatag tgatcttctg gcattttata atcaatgtca gacccgcaag agtaacgaaa 9000

agatagatat tattagcaga aatatctatc caccgcgggt tacaaaatg gatccaaag 9060
taaaattcct tcattgttat gcttgggaag aaacgggctt tccgcaaccg tggatcaatg 9120
aatttaatcg ggaacttgac ggagtgcgt gtacttcgga acatgttcgt aaaatactga 9180
ttgataacgg actgaatgtg cccgcatttgc ttgttggcaa tggctgtgac cattggctca 9240
atatcccagc cgagacgaca aaagatgtgg atcacgaaac attccgtttc ctgcacgtct 9300
cttcttgcgtt cccacgcaaa gggatacagg caatgcttca ggcttggggg aaggcggtca 9360
ctcgctgtga caatgttatac ttaatcatta agactttaa caatccgcac aatgaaattg 9420
acgcatggct ggctcaggcc caggctcaat tcatagacta tcccaaagtt gaagtgtatca 9480
aagaggatat gtcagccacc gagcttaaag ggctttatga aagctgtgat gttttgggtt 9540
ctccaggttg cgctgaaggc tttggtttac ctattgctga agcaatgctg agtgggctac 9600
cggttatcgt caccaattgg agcgggcaac ttgattttgt taattcacaa aattcatggc 9660
tggttgacta tcagttcact cggtaaaaaa cgcactttgg tctgtttcc tcagcctggg 9720
ccagtggttga tattgacaac ttaacagatg cattaaaagc ggcagcctca accgataaat 9780
cagtgtcg 10000
tgacatggcc aatgctggc gcgagttct tctgcagcag tttacctgga 9840
aagcggtggc tgatcggttct tgccaggcgg tcaagactct gcgtgcgcattt attgatattg 9900
cacagcatcg ggcgcgcatt ggctgggtga cgacctggaa cacgaaatgt gggatcgca 9960
cctattccca gcatctggtg gaaagcgcac ctcatggcgc ggatgttgc 10020
aggcagcgc tggcgatctt gtgtgtgcag acgaagagtt tgtacttcgc aactggattt 10080
taggtaaaga gagcaactat ctggaaaacc tccagccaca cattgatgct ctgagactcg 10140
atgtcattgt gatccaattc aactatggat tcttaatca tcgagaactg tcggcgat 10200
ttcgtcgcca gcatgacgcc ggtcggtcag ttgttatgac gatgcactca actgtggatc 10260
cgctggaaaa agagccgagc tggaaattcc gtcttgcgtga aatgaaagag ggcgtggcac 10320
tttgcgaccg gttgttggtg cattcgattt ccgatatgaa ccgccttaaa gattnaggct 10380
taactgcgaa tggtgcattt ttccgcacg gtgttatcaa ctactccgca gcgagcgtca 10440
cacgtcaaca gcagtcttta ccgctaatttgcgatccatgg cttctgctta ccgcataagg 10500

gcctgatgga actagtagaa tccgtccata gactcaagca agccggtaaa ccggttcggtt 10560
tacgactggt gaacgcagag tatcctgttg gggagtcacg cgatctggtg gcagagctta 10620
aagctgctgc tcagcggtaa ggtgttaccg atctgattga gatgcataat gatttcctac 10680
ctgatgcgga gagtctgcgg ttgcttcag aagccgatct tctgattttt gcttatcaga 10740
atactgggaa gtctgcttagc gggcggtac gttatggtat ggcgactcaa aaacctgttg 10800
cggttaacgcc cctggcgata tttgatgatt tggacgatgc cgtctttaaa tttgatggat 10860
gcagcgtcga tgatatcagt caggggattt accggatcct gaattccatc cgtgaacaga 10920
actcttgggc aaccaggact caacaacgtg ccgatgcattt gcgaaaacaa catgattatc 10980
aagctgtttc acgcccgtctg gttaatatgt gtcaaggctt agctaaagct aaatattttt 11040
aataaaaaata tctcttttgtt atttttgcc tttgaataca agaggggtta gataatgtgt 11100
catttattat gaaaattattt tttgctactg agccaattaa ataccattt acgggcatcg 11160
gtcggtattt cctggagctg gttaagcggc tggcggtcgc ccgcgaaattt gaagaattaa 11220
agctatttca cggtgcgtcg tttatagaac agatccctttt ggtggagaat aaaaggata 11280
ccaaagccag caatcatggt cgtctgtcg cgtttctacg ccgacagacg ctgttattttt 11340
aggcttatttca cttgctgcat ccgcggcgcc aggctgggc attgcgcgac tataaggattt 11400
atatctacca tggccccat ttttatctgc cgatccaaactt ggaacgcgcgccc gtgaccacgt 11460
ttcatgacat atccattttt acctgcccgg aatatcatcc aaaagatcgg gttcgctata 11520
tggagaagtc cctgcgtgag agtctggattt cggcaagactt gatcctgacc gtttctgattt 11580
tctcgccag tggaaattttt cgtttttca actatccggc ggagcggatc gtaaccacca 11640
agcttagcctg cagcagtgtac tataccac gcagccggc agagtgtctg ccggtaactgc 11700
agaaatatca gctggcgtgg caggcctacg cgctatataat cggcactatg gagccacgta 11760
aaaatatccg aggccctgctg catgcctatc agctgctacc gatggagatc cgcatgcgt 11820
atccgctaattt ccttagcggc tatcgccgtt ggaaagacga tgtgctgtgg cagtttagtgc 11880
agcgcggtaac tcggaaaggc tggatccggtt acctcggtt tggatccggat gaagacctgc 11940
cgtatctgttca cgcagcggcc agagtctttt tttatccctc cttctacgag ggattcggtt 12000

tacctattct tgaagcgatg tcttcgggtg tgccggtagt atgctccaat gtcaccttt 12060
tgcctgaggt tggtggcgat gccggcctcg ttgccatcc taatgatata gacgcgatta 12120
gcgcgcaaat tttgcagagc ctgcaagatg atagctggcg gaaaatcgcc accgcgcg 12180
gtcttgctca ggcgaaacag tttcgtgg agaactgtgc gacacagacc attaacgcct 12240
ataaaattact ctaagggtgt cagttgagag ttctacacgt ctataagact tactatccg 12300
atacctacgg cggtattgag caggtcattt atcagctaag tcagggctgc gcccggcggg 12360
gaatcgcagc cgatgtttc acttttagcc cggacaaaga tacaggtcct gtcgcttacg 12420
aagatcatcg ggtcatttat aataaacagc ttttgaaat tgccctcacg ccgtttcgc 12480
tgaaagcggt aaagcgaaaa aagctgatta aagatgacta cgatatcatc aactaccatt 12540
ttccgtttcc ctttatggat atgctgcattc ttccggcg gcctgacgccc aggactgtgg 12600
tgacctatca ctctgatata gtgaaacaaa aacggtaat gaagctgtac cagccgctgc 12660
aggagcgatt tctcagcggc gtagattgca tcgttgcctc gtcgcccata tacgtggctt 12720
ccagccagac cctgaaaaaaaa tatctggata aaacgggtt gatcccggtt ggtctggagc 12780
agcaggacgt gcagcacat ccgcagaggg tcgcgcactg gcgggaaact gtcggcgata 12840
agttcttctt cttcgtcggc actttccgct actacaagg gctgcattt ctgatggatg 12900
ccgctgagcg tagccgactg ccagtgggtt ttgttaggggg cggccgctg gaatcgaaag 12960
tgcggcgtga agcgcagcag cgcggcgtga gcaatgtat gtttaccggc atgctcaacg 13020
acgaagataa gtacattctc ttccagctct gccggggcgt ggtattcccc tcgcatttc 13080
gctctgagggc gtttggcatt acgttattgg aaggcgacg ctttgcagg ccgctgatct 13140
cttgcgagat cggtacaggt acctcttca ttaaccagga caaagtggatg gtttgcgtga 13200
ttccgcccggaa tgatagccag gcgcgtggagg aggcgatgaa tgagctctgg aataacgagg 13260
aaacccctccaa ccgctatggc gaaaactcgc gtcgtcggtt tgaagagatg tttactgccc 13320
accatatgat tgacgcctat gtcaatctct acactacatt gctggaaagc aaatcctgag 13380
cgccgcgag ctcgtcgact cgaggatccg tgttagggctgg agctgcttcg aagttccat 13440
actttctaga gaataggaac ttccggatag gaactaaggaa ggatattcat atggataaag 13500

ccgtaagcat ataagcatgg ataagctatt tatactttaa taagtacttt gtatacttat 13560
ttgcgaacat tccaggccgc gagcattcag cgcggtgatc acacctgaca ggagtatgta 13620
atgtccaagc aacagatcgg cgtagtcggt atggcagtga tgggacgcaa cttgcgc 13680
aacatcgaaa gccgtggta taccgtctt atttcaacc gttcccgtga gaagacggaa 13740
gaagtgattg ccgaaaatcc aggcaagaaa ctggttcctt actatacggt gaaagagttt 13800
gtcgaatctc tggaaacgcc tcgtcgcatc ctgttaatgg tggaaagcagg tgcaggc 13860
gatgctgcta ttgattccct caaaccatat ctcgataaag gagacatcat cattgatggt 13920
ggtaacacct tcttcagga cactattcgt cgtaatcgtg agcttcagc agagggctt 13980
aacttcatcg gtaccggtgt ttctggcggt gaagaggggg cgctgaaagg tccttctatt 14040
atgcctggtg gccagaaaga agcctatgaa ttggtagcac cgatcctgac caaaatcgcc 14100
gccgtagctg aagacggtga accatgcgtt acctatattg gtgccgatgg cgccagg 14160
tatgtgaaga tggttcacaa cggatttcaa tacggcgata tgccagctgat tgctgaagcc 14220
tattctctgc ttaaagggtgg cctgaacctc accaacgaag aactggcgca gaccttacc 14280
gagtggata acggtaact gagcagttac ctgatcgaca tcaccaaaga tatcttcacc 14340
aaaaaaagatg aagacggtaa ctacctggtt gatgtgatcc tggatgaagc ggctaacaaa 14400
ggtaccggta aatggaccag ccagagcgcg ctggatctcg gcgaaccgct gtcgctgatt 14460
accgagtctg tgtttgcacg ttatatctt tctctgaaag atcagcgtgt tgccgc 14520
aaagttctct ctggccgca agcacagcca gcaggcgaca aggctgagtt catcgaaaaa 14580
gttcgtcgtg cgctgtatct gggaaaatc gtttcttacg cccagggtt ctctcagctg 14640
cgtgctgcgt ctgaagagta caactggat ctgaactacg gcgaaatcgc gaagatttc 14700
cgtgctggct gcatcatccg tgcgcagtgc ctgcagaaaa tcaccgatgc ttatgccaa 14760
aatccacaga tcgctaacct gttgctggct cctgtacttca agcaaattgc cgatgactac 14820
cagcaggcgcc tgcgtgatgt cgttgcttgc acgtacaga acggatttcc gttccgacc 14880
ttctccgcag cggttgccta ttacgacagc taccgtgctg ctgttctgccc tgcgaacctg 14940
atccaggcac agcgtgacta ttttggtgcg catacttata agcgtattga taaagaaggt 15000

gtgttccata ccgaatggct ggattaa	15027
<210> 15	
<211> 11283	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> example 015 rfb locus nucleotide sequence - 015-EPA production strain stLMTB11738	
<400> 15	
atgacgaatt taaaaggcagt tattcctgta gcgggtctcg gatatgcata tttgcctgcc	60
actaaggcga tacccaaaga gatgctacca atcgtcgaca agccaatgtat tcagtgacatt	120
gttgacgaga ttgtggctgc agggatcaaa gaaatcctcc tggtaactca cgcgtccaag	180
aacgcggctcg aaaaccactt cgacacctct tatgagtttag aatcactcct tgagcagcgc	240
gtgaagcgtc aactgctggc ggaagtacag tccatctgtc cggccggcgt gaccattatg	300
aacgtgcgtc agggcgaacc ttttaggttta ggccactcca ttttgtgtgc ggcacccgtc	360
attggtgaca acccatttgt cgtggtaactg ccagacgttg tgatcgacga tgccagcgc	420
gaccgcgtac gttacaacct tgctgccatg attgcacgtt tcaacgaaac gggccgcagc	480
caggtgctgg caaaacgtat gccgggtgac ctctctgaat actccgtcat ccagactaaa	540
gagccgctgg accgtgaggg taaagtcaac cgcattgttg aatttatcga aaaaccggat	600
cagccgcaga cgctggactc agacatcatg gccgttaggtc gctatgtgct ttctggcgat	660
atttggccgg aactggaacg tactcagcct ggtgcattgg gacgtattca gctgactgat	720
gctattgccc agctggcgaa aaaacaatcc gttgatgcaa tgctgatgac cggcgcacgt	780
tacgactgctc gcaaaaaaat gggctatatg caggcggttgc tgaagtatgg cctacgcaac	840
ctgaaagaag gggcgaagtt ccgtaaaggt attgagaagc tggtaagcga ataatgaaaa	900
tctgaccgga tgtaacggtt gataagaaaa ttataacggc agtggaaatt cgcagcaaaa	960
gtaatttggtt gcgaatcttc ctggcggttgc tttatataaa ccatcagaat aacaacgagt	1020
tagcagtagg gttttattca aagtttcca ggatttcct tggccaga gcggattgg	1080

aagacaatta gcgtttgaat ttttcgggtt tagcgcgagt gggtaacgct cgtcacatca 1140
taggcattgca tgcatgtc tcgttagctgt aaagccaggg gcggtagcgt gcattaatac 1200
ctctattaat caaaactgaga gccgcttatt tcacagcatg ctctgaagta atatgaaata 1260
aattaagcta gcatgagcaa aactaaacta aatgttctt accttgcaat aagtcagggt 1320
gccaaattacc tactgccatt attaattttt ccttatcttgc tttagagtcat tggtgtatcg 1380
aattttggtg atctgagttt ttcattgata actatacaag tttgttaat gggtgttcaa 1440
tatggtttg gatatagtgg gacaagagaa atagcactaa ataacgataa aaaataccat 1500
tctgaattttt tttgcgggtgt ggtgcttgcgt cgttttatat taatgctaatt tgcagctata 1560
ataactcataa tactctgttt ttttatgttt tttaacgacg ttaagtcttt gttatgtgtt 1620
ggttttctgt ccgtaattgc aggtgttttc aatccaaatt gggttttgcga aggtaaggaa 1680
atgatgagtg tgatggctgt gctgtcacta ttttcacgag gcatacgatc cggtgcagtt 1740
tatctaatta taaaaccgc aacgcccgt tacatcagtgc ctttattttt gagcatgcca 1800
tatattttgtt attcattctg tggcggttgc tacattactta ttatcaagga gatTTTTTA 1860
tgttaggccac cgataaagaa aattcaagta attttaaaaaa atggatttca tttttttgtt 1920
tcaacacttg cgacttagtgc atacacaatg ttgacccttc ttgtattggg tggcgatct 1980
ggaaagtttg atgttaggcat ctttaactca gctaacatga tcaaacaagg tttggctgga 2040
cttgcacatc cattagtcca agcttttat ccaagaatta acattttgcga aagagagaat 2100
ccatatattt caaactttaa atctagaatg attcttaaat acttgcttgc tttttacatg 2160
gcttttagcaa taccatTTTt acttttgc aaccaattat cattattaat attcggcatg 2220
aaaggtaag taattgcagg tgcaatgcaa ttaatgacat tgcttcctat attcataggt 2280
tttaatacag ttgtcggtt acttgcattt gtacctaattt ggtatgcaaaa acagtatttc 2340
aaatctattt tccttaggaac tattacttgtt ttaagcatag tttatccagc atgtaaatatt 2400
tatggagcaa cgggtgcgtat tgtgagtctt attgttagctg aaattttcgat tggcatggaa 2460
atgcttaaac aattcattaa agtaaataaa accgtatgtt ggcctcataa attatgaaata 2520
tctcggtaat aatatctgtt tggaaacgccc cagttcaattt agaatttgattt ctctctgagc 2580

tcgattctca ggctaaagac aatagtctac acctagaagt aattgttcc gatagtcata	2640
gtggtaaaga aattgatgat gtagttgctg ataatattca taaaaagaaa aatattaata	2700
ttatccatca acatactaaa aatatactct ccgctaagcg caatttcgga gcatccctag	2760
cccatgggga ttattnata tttcttgatg atgattgtat acccgcaagt ggatatata	2820
catcgttgct gaactattta aaaaaatga atagtaaaag cgttttatgt ggggaagtta	2880
gattcgaaaa tgaactcatt gagaccagca attactatcg ctacaggaac tctttacacc	2940
ctaagtttag tgatagtcct gatatctcta tgaatgcctg gactttgtc gcaatgaatt	3000
gtgttcttga tagaaaggca ttttcatcag gtatagttc atataatgaa aattttattt	3060
gttatggttg tgaagatcat gagtttgggt ggcaacttga aaaaaatgac ttcaaaattt	3120
ttttgctga ttttaaaata ttacatcagc aatacagtgg cgatatagaa ggatatacaa	3180
aaaaaattcg tgctacagca cgtgatggta tgaatgtatt aagcaaagta aggcctgaaa	3240
tgtttctac taataaaaaaa ttattcctag ttgagaaaat atttagtaaa cacaaaacgt	3300
ttagtaaaat atgccaatca atattttca ataaattttat ttttaaaaaaa ataatacaat	3360
ttttaaaaaa aacagatgca aataaaaaac tctatttccc aattctttac agatatgtgt	3420
tgatttcggc atatatacat ggtattggag agcgtggcac ctcaaaaaca gatgatttgc	3480
ttaagaactg gtatata>tag atgatgctat cttcattttat taagacattt gtatggaagg	3540
taaaaaacaa tgaagtataa tgcattgatg gctttttat tatttttgt tgttttttt	3600
agattgtcgc tgataatacc tttcttatat ttggcattta ttcctgcatt ttttggatt	3660
atgtattttag tgcgttaattt tatgattact atgggcaatg gattggatc tatagatcgt	3720
aaaaatttgc tgctgttattc tatattcata attattttt tattttgtt ggtttcgat	3780
ttgtttcaaa aaagccattc ttttcaaagt tattttaccg ttagattatt tatgttgtt	3840
ttatttcat ttgttcctgc gtattattta gttaaatagat tcataaagggg tgacttgaaa	3900
ttaatggagc gaatatttagt gtattctctc tgggttcaaa tagttttttt ttttggat	3960
tatataagtc cagagttaaa aagattgtta tatactttct ttggatgtc tgactctgtt	4020
aatctttggg aacaaaatgc taaagtaaga ggatttgggt tgcgggtga aataaatttc	4080

atgacaccat ttttgatgt ctagatgtca ttttttatga tggaaaaggcg ttatgttta	4140
attacttaa tttgtctgac tcaaatacgta aattctaaca tggctgtgat tgcagccatt	4200
attggtatcg gttgctctag acttaatatt aatataaaaa ttgcaacagt attgatttg	4260
ggagtttag tttatagctt aggagcggtg ttcttcctc gattttatga tgagttcgaa	4320
tctggagatg gcacaagaac tctggatatc ttattacagc aacatgtgtt tggttaggt	4380
aatttagatt ttttaatat tatatttggaa ttacagcaaa acatatcttc atcaatcccc	4440
gatattaaac aaagttcgga tatgggctgg gttatactgt ttaattacgg tgggttaaca	4500
tttattacac tcttttatt ttaatcttt actattctta ttgcgacatt tggaatgaca	4560
tatcaagcaa ttatatggat gttaattggg ataatttca ataccaaagg ttttagttt	4620
ggatctaacg gctatttcctt tctatcttt atatataatgt tttgaatag agtaacactt	4680
agtggacaga gttcaattac taataagtta ggtcaagtaa gtaaatagct tccagagtg	4740
atttgtcaat gatttgaggt tcggttatta tgtttcatc taaaacactg ttaattactg	4800
gtggactgg ctcttcggg aatgctgtat taaatagatt tcttgataca gatattgcag	4860
aaatccgtat attagtcgt gataaaaaaaaa aacaagatga tatgcggaaa aaatacaata	4920
atcaaaaatt aaagttctat attggtgatg tcagagatta ccgtagtatt ttgaatgcga	4980
ctcgccgtgt tgattttata tatcatgcag cggcacttaa gcaagttcca tcatgtgaat	5040
ttcatcctat ggaagccgtt aaaactaata tccttggtac ggaaatgtt cttgaagcag	5100
ctatagcgaa tgaagtgaag agggttgtat gcctaagtgac tgataaagct gtataccgaa	5160
ttaacgcaat gggattttca aaagctatga tggaaaaggat catggtcgca aaatcccgt	5220
atgttgatcg caataaaaca gtaatatgtg gtacccgtta tggaaatgtt atggcatctc	5280
gcggttcagt tattccatta tttgttgatc ttattagagc gggcaagcca ctcacaataa	5340
ctgatcctaa tatgaccgc tttatgtatgc ctcttgagga tgccgttagat ttatcttt	5400
atgcgtttga acatggtaat aatggtgata tctttgtgca aaaaggcacct gcagcaacta	5460
ttgacacatt agctattgct ttaaaggaat tactaaatgt tcctgaccat ccggtaaatg	5520
tcattggAAC gcgtcatggc gagaaattat atgaagctt acttagtcgt gagaaatga	5580

tcgctgctat agatatggc gattattacc gtgtccgc agatttcg gaccttaatt	5640
atggcaaata tggtagcaaa ggtgatagcc gaatatctga aatagaagat tataactctc	5700
ataatactca acggtagat gttgaaggca taaaagagct ctgtctaaaa ttgcctta	5760
ttcgagcaat tcgtgctggt gaaaaatata atctggattc atgatatgaa aatattagtt	5820
actggtgcaa atggtttat tggtcgtaat ttatgttga ggcttgagga acttggttat	5880
aaagatctta tttagaattga tcgagaatca acgaagcaag atcttgaaca aggcttacag	5940
gatgccgatt ttatccatca cttagctggt atcaatagac ctaagactga tgatgagttt	6000
atttctggaa acagtgattt aacaaagcat atagttgagt atctccccc tattggtaag	6060
aatacaccaa ttatgctaag ttctcgata caagctgaac ttaataatgc ttatggggtt	6120
agcaaagctg tagctgaaag ctatgtcgaa aaatatgctg ctgctagtgg ttcttcgtat	6180
tatattttca gatatccaaa cgtttttgtt aaatgggtta agccaaacta taatttttt	6240
atagcaactt ttgctacaa tattccat gatattgaga ttactatcaa tgatgcagca	6300
gcccggcgtca atctggtcta tattgatgat gtttgtactg atgctatagc tcttcctct	6360
gggacgggtt aaagtggata taaagttgtt gcaccaattt attcaacaac agttggtaaa	6420
gttgcagaat taatttatac cttccaaaat agccgttcca ccctgatcac agaggctgtc	6480
ggggcgggat ttaccggcgtgc attgtattct acatggctga gttatcc accagagaag	6540
tttgcgtaca aggtaccc ttatgggat gcccggag tctttgtga gatgtgaaa	6600
acgccttcag cggggcgtttt ttcattttt actgctcacc ctggattac gcgtggcgg	6660
cattaccatc acagtaaaaa tgagaagttt ttggcattc gaggtcaggc atgcttaaa	6720
tttgaacatg tgattaccgg tgagcgatg gaaatgaaag tttcatcggg tgagttaag	6780
attgttggaa cagttccctgg ttggacacat gacattacaa atattggaaac tgatgaatta	6840
atagtcatgc tctggcaaa taaaatttc aaccgtgtatg agcccgatac tattgcgaga	6900
cctctataat gaaaaattaa aaagttatgt ctgttggg aaccggcct gagattatcc	6960
gtttgtcgag gtttcttgct aagttgtatg aatactgcga gcatattatt gtccatactg	7020
gtcaaaattaa tgattacgaa taaaatgaaag tggcttcaa tgacttgggt gttcgaaaac	7080

ctgattattt tttaaatgca gcgggtaaaa atgcggcgga aaccattggc caggttatta	7140
ttaaggtaga tgaagtatta gaaatcgaaa aacctgaagc aatactggta ttggggcata	7200
cgaattcatg tatttctgcc attccggcca aacgccgtaa agtcgcctata tttcatatgg	7260
aagcaggtaa ccgttgttgc gatcaacgac tgccctgaaga aaccaacaga cgtattgttgc	7320
accatacggc tgatatcaat atgacctaca gtgatattgc tcgtgaatat ctcttggtcg	7380
aaggtatccc agctgatcgg atcataaaaaa ctggtagccc tatgttttag gttcttcat	7440
attatatgcc ccaaattgtat ggttcagatg tgctatcgac tttgaatcta cagtctggtg	7500
agtttttgt agtaagtgcg catcgtaag agaatgttga ttctccaaaa cagctcgtaa	7560
agcttgcgaa cattctaaat actgttgctg aaaaatataa tcttccagtt attgtctcca	7620
cacacccaag gacacgtaac cgaatccgtg agcaaggaat tgaatttcat tcaaataaa	7680
atctactgaa accattgggt ttccatgatt ataaccactt gcagaagaac tcacgagctg	7740
tgcttcaga tagcggtact atcactgaag agtcatccat catgaatttc ccagcggtaa	7800
acatccggga agcgcgttag cgccggaaag gctttgagga agcatccgtc atgatggtg	7860
ggtagatgt tgaacgcgta ttacaagcgc tggatattct ggcaacacaa ccgcgagggtg	7920
aagtccgtct ttacgtcag gttatgttattt acagcatgcc aaatgtgtcg gataaagttg	7980
tcagaattgt tcactcttac acagattatg ttaagagatg cgtctggaaa gaatattgtat	8040
gaaacttgct ttaatcatag atgattacct gcccaacagt actcggttg gtgcacaaat	8100
gtttcatgaa cttgctcaag aatttatcca gcgtggcac gatgttacgg taattactcc	8160
tggtagggc atgcaagaag agatttctt tgataccctt cagggggtaa aaacatggcg	8220
ttttaaaagc gggccgctca aggatgtttag taaaatttcg cgagcggtca atgaaacgct	8280
tttgcctat cggcggttga aagccatcaa aaaatggta aaaaaagaga cctttgaggg	8340
ggtagttat tattcacctt ccatattctg ggggcctta gttaaaaaaaaa ttaaagctcg	8400
ttgccaatgt cctgcttac ttatTTTaa agatatgttt ccacaatggg taattgtgc	8460
aggaatgctt aatgctggtt cccaaataga acgctacttt cgtcttttg aaaaatatc	8520
ttatcgtagt gcaaattcgta ttggacttat gtctgataag aatcttgcgt ttttcggaa	8580

agataataaa ggctatccgt gcgaagttt gcgtaattgg gcatccctaa caccaacgt 8640
catacccaag gattatatac cactacgtaa gcgacttggc ctagaggata aaaccatttt 8700
cttctatggt ggaaacatag gtcatgcaca ggacatgaca aacttgatgc gacttgtgag 8760
aaacatggca gcatatcctc aagctcattt cctatttatt ggccaggggg atgaagttga 8820
attaattaat tcattagcat ctgagtggc attgacgaat ttcacctatt tgccctcggt 8880
taaccaagat gaatttaagt tcattttgtc ggaaatggat atcggcttgt tttctttc 8940
cgctagacac tcttcccata atttcctgg taagttatta ggctatatgg ttcagtcgct 9000
acctatttta ggttagcgtaa atgccggaaa tgatttgctc gacattgtca atcaaaataa 9060
tgcgggatta atccatgtca atggtgagga cgataaatta tgtcaatctg cgctattaat 9120
gttgcatgat attgatgtgc gccggcaact tgggtcgaaa gccaatataat tggtaaaga 9180
acaattctcc gttgagtctg cggcacagac gatagaaatg aggttggagg catgcaatgc 9240
gattaattga taatgaccaa ctcgacgaat tatatgatca agccggcaa tcggaacggtt 9300
tacgttccca ccttatgatg cacggctcgc atcaagaaaa ggtacagcgt ttacttattt 9360
cattagtaaa gggcagctat gttgaaccgc attatcacga acttcctcat cagtggaaa 9420
tgttcattgt tatggagggg caacttcagg tttgtttgtt tggttagaaat ggtgaggtta 9480
taaagcaatt tatagcagga gataatactg gaatgagcat tgtggagttt tctccggcg 9540
atatacacag tgtcgaatgc ctatctccgc gtgctttat ggtggaaat aaggaggggc 9600
catttgaccc ttctttgca aaatcggtcg tgtgaggccc cggcagctcg tcgactcgag 9660
gatccgtgtta ggctggagct gcttcgaagt tcctatactt tcttagagaat aggaacttcg 9720
gaataggaac taaggaggat attcatatgg ataaagccgt aagcatataa gcatggataa 9780
gctatttata cttaataag tactttgtat acttatttgc gaacattcca ggccgcgagc 9840
attcagcgcg gtgatcacac ctgacaggag tatgtaatgt ccaagcaaca gatcggcgta 9900
gtcggtatgg cagtgtatgg acgcaacctt gcgctcaaca tcgaaagccg tggttatacc 9960
gtctctattt tcaaccgttc ccgtgagaag acggaagaag tgattgccga aaatccaggc 10020
aagaaactgg ttccttacta tacggtgaaa gagttgtcg aatctctgga aacgcctcg 10080

cgcacccgt taatggtaa agcaggtgca ggcacggatg ctgctattga ttccctcaa	10140
ccatatctcg ataaaggaga catcatcatt gatggtggtt acaccttctt ccaggacact	10200
attcgctgtc atcgtgagct ttcagcagag ggctttaact tcacgtgtac cggtttct	10260
ggcggtaag agggggcgct gaaaggtcct tctattatgc ctggtgccca gaaagaagcc	10320
tatgaattgg tagcaccat cctgacccaaa atcgccgccc tagctgaaga cggtgaacca	10380
tgcgttacct atattggtgc cgatggcgca ggtcactatg tgaagatggt tcacaacgg	10440
attgaatacg gcgatatgca gctgattgct gaagcctatt ctctgcttaa aggtggcctg	10500
aacccacca acgaagaact ggcgcagacc ttaccgagt ggaataacgg tgaactgagc	10560
agttacctga tcgacatcac caaagatatc ttccaaaaa aagatgaaga cggttaactac	10620
ctgggtgatg tgatcctgga tgaagcggtt aacaaaggta ccggtaaatg gaccagccag	10680
agcgcgctgg atctcggcga accgctgtcg ctgattaccg agtctgtgtt tgcacgttat	10740
atctcttc tgaaagatca gcgtttgcc gcatctaaag ttctctctgg tccgcaagca	10800
cagccagcag ggcacaaggc tgagttcatc gaaaaagttc gtcgtgcgt gtatctggc	10860
aaaatcgaaa cttacgccccca gggcttctct cagctgcgtg ctgcgtctga agagtacaac	10920
tgggatctga actacggcga aatcgcaag attttccgtg ctggctgcat catccgtcgc	10980
cagttcctgc agaaaatcac cgatgcttat gccgaaaatc cacagatcgc taacctgttgc	11040
ctggctccgt acttcaagca aattgccat gactaccagc aggctgcgtg tgcctattac	11100
gacagctacc gtgctgctgt tctgcctgcg aacctgatcc aggcacagcg tgactat	11220
ggtgcgccata cttataagcg tattgataaa gaaggtgtgt tccataccga atggctggat	11280
taa	11283

<210> 16
<211> 13435
<212> DNA
<213> Artificial Sequence

<220>
<223> example 016 rfb locus nucleotide sequence - 016-EPA production

strain stLMTB11739

<400>	16								
atgacgaatt	taaaaggcagt	tattcctgta	gcgggtctcg	ggatgcata	gttgcc	60			
actaaggcga	tacccaaaga	gatgctacca	atcg	tcgaca	agccaat	tgat	tcatt	120	
gtt	gacgaga	tttgtggctgc	aggatcaaa	gaaatc	cctcc	tgt	aactca	cgcgttcaag	180
aacgcgg	gtcg	aaaaccactt	cgacac	ctct	tatgagtt	tag	aatcactc	cct tgagc	240
gtga	agcg	tc	aactg	ctggc	ggaagt	acag	tccat	ctgtc cgc	300
aacgt	gcgtc	aggc	gaacc	tttaggtt	ta	ggccact	cca	ttttgtgtgc	360
att	gggt	gaca	acc	cattt	gt	cgtt	ccagac	gtt g	420
gaccc	gctac	gtt	aca	acct	tg	ctgcat	att	tgatcgac	480
cagg	tgctgg	caa	aaacgtat	gccgggt	gac	ctct	ctgaaat	actccgt	540
gagcc	gctgg	acc	gtgaggg	taa	agt	cagc	cgcat	ttttatc	600
cagcc	cgaga	cg	ctggactc	agacat	catg	gccgt	taggtc	gctatgt	660
attt	ggccgg	aact	ggaacg	tact	cagc	ct	gtgg	gacgtatt	720
gctatt	gccc	cg	actggcgaa	aaa	acaat	cc	gtt	gatg	780
tac	gact	g	caaaaaaaat	gg	ctat	at	ggctt	ta	840
ctgaa	agaag	gg	ggcgaagtt	cc	gttaa	agg	tt	tttgc	900
tctgacc	gga	tg	taacgg	tt	ataac	ggc	tttgc	tttgc	960
gtaattt	gtt	gc	aatcttc	ct	gccc	tttgc	tttataaa	ccatc	1020
tagc	atgtt	tttatt	ca	aagttt	cc	ggat	tttcc	caga	1080
aagaca	atta	gc	gtt	tttgc	tttgc	tttgc	tttgc	tttgc	1140
taggc	atgca	tg	cagt	gtc	tgg	tag	ctgt	aaagg	1200
ctct	ttaat	ca	aaact	gaga	cc	gc	ctt	tttgc	1260
aatta	agtga	aa	ataact	ttgt	tttgc	tttgc	tttgc	tttgc	1320
cacattataa	ataata	ca	gca	ggat	ttgtt	ttgtt	ttgtt	ttgtt	1380
aaccgg	aat	cact	tgta	ttttct	gtt	ttttct	ttttct	ttttct	1440

atttgcgatg cacctgcaat ggcacggatt tttgctcagc atcagccgga tgcagtgtat	1500
cacctggctg ctgaaagcca tgttgaccgt tcaattacag gccctgcggc atttattgaa	1560
accaatattg ttggtaactta tgtcctttg gaagccgctc gcaattactg gtctgcttt	1620
gatagcgaca agaaaaatacg cttccgtttt catcatattt ctactgacga agtctatggt	1680
gatttgcctc atccagatga agtaaataat acagaagaat tacccttatt tactgagacg	1740
acagcttacg cgccaaggcag cccttattcc gcatccaaag catccagcga tcatttagtc	1800
cgcgcgtgga aacgtacata tggttaccg acaattgtga ctaattgctc gaacaactat	1860
ggtccttatac atttcccggaa aaagcttatt ccactggta ttcttaatgc actggaaggt	1920
aaggcattac ctatttatgg caaaggagat cagatccgctc actgggtgttga tggtgaagat	1980
catgcgcgtg cgttatatac cgtcgtaacc gaaggtaaag cgggtgaaac ttataacatt	2040
ggtgggcaca acgaaaagaa aaacatcgat gtagtgctca ctattgtga tttgctggat	2100
gagattgtac cgaaagagaa atcttacgt gagcaaatca cttatgttgc tgatcgtccg	2160
ggcacacgatc gcccgtatgc tattgatgct gagaagattt gtcgcccatt gggatggaaa	2220
ccacaggaaa cgttgagag cgggattcgt aaaacggtgg aatggtacct gtcctaataca	2280
aatgggttg ataatgtgaa aagtggtgcc tatcaatcgt ggattgaaca gaactatgag	2340
ggccgccagt aatgaatatac ctccttttgc gcaaaacagg gcaggttagt tgggaactac	2400
agcgtgctct ggcacctttg ggtaatttga ttgctttga tggactct actgattatt	2460
gcggtgattt tagtaatcct gaaggtgttag ctgaaaccgt aagaagcatt cggccggata	2520
ttattgtcaa tgcagccgt cacaccgcag tagacaaagc agaattcagaa ccggagtttgc	2580
cacaattaat taacgcaaca agtgcgaaag cgattgcgaa agcagcaaataat gaagttggag	2640
cctgggttat ccattactcg actgattacg tcttccctgg aaatggcgat atgccccatggc	2700
tggagacgga tgcaaccgca ccactaaatg tttacgggtga aaccaagtttta gcccggagaaa	2760
aagcggttaca ggaatattgc gcgaagcattc ttatcccg gaccagctgg gtctatgcag	2820
gaaaaggaaa taacttcgccc aaaacgatgt tacgtctggc aaaagagcgt gaagaatttag	2880
cggttattaa cgatcagttt ggtgcgcggaa caggtgctga actgctggct gattgtacag	2940

cacatgccat tcgtgtcgca ctgaataaac cgatgtcgc aggcttgc acatggtag 3000
ccagtggta cacaacctgg tacgattatg ctgcgctgg ttttgaagag gcgcgaaag 3060
caggcattcc cttgcactc aacaagctca acgcagtacc aacaacagcc tatcctacac 3120
cagctcgta tccacataac tctgcctta atacagaaaa atttcagcag aactttgcgc 3180
ttgtcttgcc tgactggcag gttggcgtga aacgaatgct caatgaatta tttacgacta 3240
cagcaattta atagttttg catcttgttc gtatggtgg agcaagatga attaaaagga 3300
atgatgaaat gaaaatgcgt aaaggtatta ttttagcggg tggttctggt acacgtctt 3360
atcctgtgac tatggctgtc agtaaacagc tattacctat ttatgataaa ccgatgatct 3420
attacccgct ctctacactg atgttggcgg gtattcgcga tattttgatt atcagtagcac 3480
ctcaggatac tcctcgaaaa caacaattgc tgggtgacgg tagccagtgg ggcctgaatc 3540
ttcagtagaa agtgcacccct agccagatg gcctcgcgcgca ggcatttatac atcggtaag 3600
agtttattgg tggtgatgat tgtgctttgg ttcttggta taatatctt tacggtcacg 3660
atctgccaa gctaattggag gccgctgtta acaaagaag tggtgcaacg gtattgcct 3720
atcacgttaa tgatccagaa cgctatggtg tcgttgagtt tgataaaaac ggtacggcaa 3780
tcagtcgaa agaaaaaccg ttagaaccaa agagtaatta cggcgttaca ggtctgtact 3840
tttatgataa cgacgtggtt cagatggcga aaaacttgcgaa gccgtctgca cgtggtgagt 3900
tagaaattac agatattaac cgtatttatac ttgagcaggg acgtctgtct gtcgcgtatga 3960
tggggcgtgg ctacgcgtgg ctggacacgg ggactcatca gagtctgata gaagcaagta 4020
attttattgc gacaatttgcgaa gagcccccagg gattgaaggt ttccctgtcct gaagagattg 4080
catttcgtaa aggtttattt gatgttgagc aagtaagaaa attagctgtta ccactaataa 4140
agaataatta tgggcagttat ctttataaaa tgacgaagga ttcaattaa tgaatgtgat 4200
tagaactgaa attgaagatg tgctaattct ggagccaaga gtatttggtg atgatagagg 4260
tttctttat gagagcttta atcaatcagc atttgaacat attctaggct atccggcgtcag 4320
ctttgttcaa gacaatcact cacgttcatc aaaaaatgtta ctcagaggcc ttcactttca 4380
acgcggcgag tacgcacaag ataaacttgt acgctgcact catggagcag tttttgatgt 4440

tgctgttcat attcgaccca attcggtatc ctttggtaaa tgggttggtg ttctgcttc 4500
agctgataat aagcagcagt tgtggatacc aaaagggttt gctcatggct ttttggttct 4560
gtctgatatac gctgaatttc aatataaaac tacaaactat tatcatcctg aaagcgattg 4620
tggaatatgt tggaatgatg aacgcattgc aattgattgg ccccaaacat cagggtaat 4680
ccttcgccca aaagatgaaa ggctcttac gtttagatgag cttatcagat taaaattaat 4740
tgcatgaata cgaataaatt atcttaaga agaaacgtta tataatctggc tgctcgttcaa 4800
ggtagcaatt atctttacc attgcttaca tttccatatac ttgtaagaac acttggcct 4860
gaaaatttcg gtatattcgg ttttgccaa gcgactatgc tataatgtat aatgtttgtt 4920
gaatatggtt tcaatctcac agcaactcag agtattgcca aagcagcaga tagtaaagat 4980
aaagtaacgt ctatTTTtgc ggcggtgata tttcaaaaa tagttcttat cgtcattaca 5040
ttgattttct taacgtcgat gaccttgctt gttcctgaat ataacaagca tgccgtatt 5100
atatggcgt ttgttcctgc attagtcggg aatttaatct accctatctg gctgtttcag 5160
ggaaaaagaaa aaatgaaatg gctgacttta agtagtattt tatcccgtt ggctattatc 5220
cctctaacat ttatTTTgtt gaacacaaag tcagatatacg caattgccgg ttttattcag 5280
tcaagtgcaa atctggttgc tggaattatt gcactagcta tcgttggca tgaagggtgg 5340
attggtaaag ttacgctatc attacataat gtgcgtcgat cttagcaga cggtttcat 5400
gtttttattt ccacatctgc tattagttt tattctacgg gaatagttat tatcctggga 5460
tttatatctg gaccaacgtc cgtaggaaat ttatgcgg ccaataactat aagaaacgca 5520
cttcaagggc tattaaatcc tatcacccaa gcaatatacc caagaatatac aagtacgctt 5580
gttcttaatc gtgtgaaggg tgtgatttttta attaaaaaat cattgacctg cttgagttt 5640
attgggttgtt ctttttcatt aattctgctc ttgggtgcatt ctatactagt aaaaataagt 5700
atagggccgg gatatgataa tgcgtgttattt gtcgtatga ttatatcgcc tctgcctttt 5760
cttatttcat taagtaatgt ctatggcatt caagttatgc tgaccataa ttataagaaa 5820
gaattcagta agattttat cgtgcgggt ttgttggat ttgttggat ttttccgcta 5880
acaactcttt ttaaagagat tgggtgcagca ataacattgc ttgcaacaga gtgcttagtt 5940

acgtcactca tgctgatgtt cgtaagaaat aataaattac tggtttgctg aggattttat 6000
gtacgattat atcattgttgcgtt gttctggttt gtttggtgcc gtttgtgcga atgagttaaa 6060
aaagctaaac aaaaaagttt tagtgattga gaaaagaaat catatcggtg gaaatgcgta 6120
cacagaggac tgtgagggta tccagattca taaatatggt gcacatattt ttcataccaa 6180
tgataaatat atatgggatt acgttaatga ttttagtagaa tttaatcggtt ttactaattc 6240
tccactggcg atttataaag acaaattatt caaccccttctt ttaatatga atactttcca 6300
ccaaatgtgg ggagttaaag atcctaaga agctcaaaat atcattaatg ctcagaaaaa 6360
aaagtatggt gacaaggtac ctgaaaattt ggaggagcag gcgatttcat tagttgggaa 6420
ggacttatac caagcattga taaagggtta tacggagaag cagtggggaa gaagtgc当地 6480
agaattgcct gcatttatta ttaagcgaat cccagtgaga tttacggtt ataacaatta 6540
ttttccgat cgctatcaag gtattccggtt gggaggctac actaagcttta ttgaaaaat 6600
gcttgaaggt gtggacgtaa aattaggcat tgatttttg aaagacaaag attctctagc 6660
gagtaaagcc catagaatca tctacactgg acccattgtat cagttactcg actatagg 6720
tggagcgtta gaatatcgct cttaaaatt tgagacggaa cccatgaat ttccaaactt 6780
ccaaaggaaat gcagtaataa atttcactga tgctaattgtt ccatatacca gaataattga 6840
gcataaacat tttgactatg ttgagacaaa gcatacggtt gttacaaaag aatatccatt 6900
agagtggaaa gttggcgacg aaccctacta tccagttat gataataaaa acatggagct 6960
ttttaagaaa tatagagagt tagcttagcag agaagacaag gttatatttgcggcggtt 7020
ggcccgagttt aaatattatg atatgcata agtgatatct gccgctttt atcaagtga 7080
aaatataatg agtacggatt aatgatctat cttgttaattt gtgtcttctt cattacagca 7140
tttatctgtt tatactttaa gaaggatata ttttatccag ccgtatgcgt taatatcatc 7200
ttcgcaactgg tcttattggg atatgaaata acgtcagata tatatgcctt tcagttaaat 7260
gacgctacgt tgattttctt actttgcaat gtttgcacat ttaccctgtc atgtttatttgc 7320
acggaaagtg tattagatct aaatatcaga aaagtcaata atgctatttta tagcatacca 7380
tcgaagaaag tgcataatgtt aggcttgcgtt gttatttctt ttgcgtatgtt atatatatgc 7440

atgaggtaa gtaactacca gttcgggact agcttactta gctatatgaa tttgataaga 7500
gatgctgatg ttgaagacac atcaagaaat ttctcagcat acatgcagcc aatcattcta 7560
actactttg ctttatttat ttggctaaa aaatttacta atacaaggta aagtaaaaca 7620
tttactttac ttgttttat tgtattcatc tttgcaatta tactgaatac tggttaagcaa 7680
attgtctta tggttatcat ctcttatgca ttcatcgtag gtgttaatag agtaaaacat 7740
tatgtttatc ttattacagc tgttaggtgtt ctattctcct tgtatatgct cttttacgt 7800
ggactgcctg gggggatggc atattatcta tccatgtatt tggtcagccc tataatcg 7860
tttcaggagt tttatttca gcaagtatct aactctgcc 8 a gttctcatgt cttttgg 8 ttt
tttggaaaggc ttaggggct attaacaggta ggagtctcta tgtcgttgca taaagaattt 7980
gtgtgggtgg gtttgccaac aaatgtttat actgctttt cggattatgt ttatattcc 8040
gcggagctaa gctattgat gatggttatt catggctgta tttcagggtgt tttatggaga 8100
ttgtctcgaa attacatatc tgtaaaata ttttattcat atttattta tacctttct 8160
ttcattttt atcatgaaag cttcatgact aatattagca gttggataca aataactctt 8220
tgtatcatag tattctctca atttcttaag gcccagaaaa taaagtggaaa atgtatTTT 8280
tgaatgattt aaatttctct agacgcgatg ctggattaa agcaagaaaa gatgcactgg 8340
acattgcttc agattatgaa aacatttctg ttgttaacat tcctctatgg ggtggagtag 8400
tccagagaat tattagttct gttaagctta gtacatttct ctgcggctt gaaaataaag 8460
atgttttaat ttcaatttc ccgatggcca aaccatttg gcatatatttgcattttc 8520
accgccttct aaaattttaga atagtacctc tgattcatga tattgatgaa ttaagaggag 8580
gagggggtag tgattctgtg cggcttgcta cctgtgatgt ggtcataagt cacaatccac 8640
aaatgacaaa gtaccttagt aaatataatgt ctcaggataa aatcaaagac ataaaaat 8700
ttgattacct cgtctcatct gatgtggagc atcgagatgt tacggataag caacgagg 8760
tcatatatgc tggcaacctt tctaggcata aatgttctt catatatact gaaggatg 8820
atTTTactct ctttgggtgc aactatgaaa ataaagataa tcctaaatat cttggaaat 8880
ttgatgctca atctccggaa aagattaacc tcccaggcat gcaatttgg 8940

atggagattc tgtcgaaacc tgttagtggtg cctttggcga ctattnaaag tttataacc 9000
ctcataagac atctctttat ctttcaatgg aacttccagt atttatatgg gataaagccg 9060
cccttgcgga tttcattgta gataatagaa taggatatgc agtgggatca atcaaagaaa 9120
tgcaagagat tggtgactcc atgacaatag aaacttataa gcaaatttagt gagaatacaa 9180
aaattatttc tcagaaaatt cgaacaggaa gttacttcag ggatgttctt gaagaggtga 9240
tcgatgatct taaaactcgc taaacgatat ggtctctgtg gtttattcg gcttggtaga 9300
gatgtcttat tgactcgtgt atttaccgg aactgtagaa ttattcgatt tccctgctat 9360
attcgcaatg atggtagcat taatttttgtt gaaaatttca caagtggagt cggtctcagg 9420
ctggatgcat ttggacgtgg cgtgattttt tttccgata atgtgcaagt taacgactat 9480
gttcatatcg cctcaattga gagcgttacg ataggtcggg atacgcttat tgcaagtaaa 9540
gtatttatta ccgatcataa tcacggttcc ttaaggact ctgatccaat gagttcgcca 9600
aatatacctc cagacatgcg cacgttggaa tcttcagctg ttgttaattgg ccagagggtt 9660
tggttgggtg agaatgtgac ggtttgcct ggaacaatta ttggtaatgg agtcgtagtc 9720
ggcgccaatt ctgttgttag aggttctatt cccgaaaata ctgtcattgc gggagtagcca 9780
gcaaaaaatca taaagaaata caatcatgag accaaattat gggaaaaagc atagtcgttg 9840
tttctgcggt caattttacc actggcggtc catttaccat tttgaaaaaaaaa tttttggcag 9900
caactaataa taaagaaaat gtcagttta tcgcattagt ccattctgct aaagagttaa 9960
aagaaaagtta tccatgggtt aaattcatttga agtttcctga ggttaaagggg tcgtggctaa 10020
aacgttgca ctgttgaatat gtatgttgcataaaaactttc aaaagagctg aatgctacgc 10080
attggatttg tctgcattat attacggcca atgtcgtcac taaaaaaaaga tatgtgtatt 10140
gtcataaccc tgccccttt tataaaggaa ttttattccg tggaaattctt atggaggccta 10200
gctttttctt atttaaaatg ctatacgggc tgatataaa aataaaacatt aaaaaaaaaata 10260
ctgcagtgtt tggtcaacaa ttctggatga aagaaaaattt tatcaagaaa tattctataa 10320
ataacatcat tgtcagtcgg ccagaaaattt aattatctga taaaagccaa cttactgtat 10380
atgattctca atttaagaat aacccttctg agttgacaat attttaccct gctgttccac 10440

gagttttaaa aaattacgag cttattatta gtgcagcaag gaaattgaaa gaacaatcca 10500
atattaaatt tctgcttact atcagtggta cagaaaatgc gtatgaaaa tatattatca 10560
gtcttcaga aggactggat aatgttcatt tcctcggtt cttggataaa gaaaaaatcg 10620
atcattgtta taatattca gatatagttt gtttccctc tagtttagaa acatggggat 10680
tgccgttgtc tgaggctaaa gagcgaggta agtgggtatt agcatcagat ttcccattha 10740
ctagagaaac tcttggtagt tatgaaaaga aagcttttt tgattctaatt aacgatgaca 10800
tgtagttaa acttattatt gacttcaaaa aaggtaacct caaaaaagat atctctgatg 10860
caaatttcat ttatcgtaat gaaaatgtat tagttgggtt tgatgaacta gttaaattta 10920
ttactgaaga acattgaaat ggtatataataatcgaaa cccacggaca tgaagactac 10980
atcaaaaaat tactcgaaaa tcttaatgct gacgatgagc actacaagat tatcgtaacgc 11040
gacaacaaag actctctatt attgaaacaa atatgccagc attatgcagg cctggactat 11100
attagggag gtgtatacgg ct当地ggcat aataataata ttgcggggc gtatgtaaag 11160
aaaaatata gacccgcaga tgaatgattac attttgtttt tgaatcccga tatcatcatg 11220
aagcatgatg atttgctgac atatattaaa tatgtcgaaa gtaagcgta tgcttttagt 11280
acattatgcc tggccgaga tgaagcgaaa tctttacatg attattccgt aagaaaattt 11340
cctgtgcttt ctgattttat tgtgtcattt atgttagggta ttaataaaac aaaaattcct 11400
aaagaaagta tctattctga tacgggtgtt gattggtgcg caggatcatt tatgctggta 11460
cggtttcag attttgtcg tggaaatggc ttgcgtcaag gttactttat gtactgtgaa 11520
gatattgacc tggcttgag gcttagcctg gctgggtgtca gacttcattt tggtcccgct 11580
tttcatgcga tacattatgc tcatcatgac aatcgaaattt tttttcaaa agccttcaga 11640
tggcactaa aaagtacttt tagatattta gccagaaaac gtatattatc aaatcgcaac 11700
tttgcgttcaattttcatcagt tttcaccccg taagagctcg gtacccgggc ctaggggtgt 11760
ggctggagct gcttcgaagt tcctatactt tctagagaat aggaacttcg gaataggaac 11820
taaggaggat attcatatcc gtcgacggcg gccgcctgc aggcattca gcttgcattca 11880
tatggatcgc tagcttaattt aaataaagcc gtaagcatat aagcatggat aagctattta 11940

tactttaata agtactttgt atacttattt gcgaacattc caggccgcga gcattcagcg 12000
cggtgatcac acctgacagg agtatgtaat gtccaagcaa cagatggcg tagtcggtat 12060
ggcagtgatg ggacgcaacc ttgcgctcaa catcgaaagc cgtggttata ccgtctctat 12120
tttcaaccgt tcccgtgaga agacggaaga agtgattgcc gaaaatccag gcaagaaaact 12180
ggttccttac tatacggtga aagagtttgt cgaatctctg gaaacgcctc gtcgcaccc 12240
gttaatggtg aaagcaggtg caggcacgga tgctgctatt gattccctca aaccatatct 12300
cgataaagga gacatcatca ttgatggtgg taacaccttc ttccaggaca ctattcgtcg 12360
taatcgtgag ctttcagcag agggctttaa cttcatcggt acgggtgttt ctggcggtga 12420
agagggggcg ctgaaaggc cttctattat gcctggtggc cagaaagaag cctatgaatt 12480
ggtagcaccg atcctgacca aaatcgccgc chtagctgaa gacggtaac catgcgttac 12540
ctatattggc gccgatggcg caggtcacta tgtgaagatg gttcacaacg gtattgaata 12600
cggcgatatg cagctgattt ctgaagccta ttctctgctt aaaggtggcc tgaacctcac 12660
caacgaagaa ctggcgcaga cctttaccga gtggaaataac ggtgaactga gcagttacct 12720
gatcgacatc accaaagata tcttcaccaa aaaagatgaa gacggtaact acctgggtga 12780
tgtgatcctg gatgaagcgg ctaacaaagg tacggtaaa tggaccagcc agagcgcgct 12840
ggatctcggc gaaccgctgt cgctgattac cgagtctgtg tttgcacgtt atatctttc 12900
tctgaaagat cagcgtgttgc cgcacatctaa agttctctct ggtccgcaag cacagccagc 12960
aggcgacaag gctgagttca tcgaaaaagt tcgtcgtgcg ctgtatctgg gcaaaatcgt 13020
ttcttacgcc cagggcttct ctcagctgcg tgctgcgtct gaagagtaca actggatct 13080
gaactacggc gaaatcgca agattttccg tgctggctgc atcatccgtg cgcagttcct 13140
gcaaaaaatc accgatgctt atgccaaaaa tccacagatc gctaacctgt tgctggctcc 13200
gtacttcaag caaattgccc atgactacca gcaggcgctg cgtgatgtcg ttgcttatgc 13260
agtacagaac ggtattccgg ttccgacctt ctccgcagcg gttgcctatt acgacagcta 13320
ccgtgctgct gttctgcctg cgaacctgat ccaggcacag cgtgactatt ttgggtgcga 13380
tacttataag cgtattgata aagaaggtgt gttccataacc gaatggctgg attaa 13435

<210> 17
<211> 13228
<212> DNA
<213> Artificial Sequence

<220>
<223> example 018A rfb locus nucleotide sequence - 018A-EPA production
strain BVEC-L-00559

<400> 17
atgacgaatt taaaaggcagt tattcctgtta gcgggtctcg ggatgcataat gttgcctgcc 60
actaaggcga tacccaaaga gatgctacca atcgtcgaca agccaatgtat tcagtcatt 120
gttgcacgaga ttgtggctgc agggatcaaa gaaatcctcc tggtaactca cgcgtccaag 180
aacgcggctcg aaaaccacctt cgacacctct tatgagtttag aatcactcct tgagcagcgc 240
gtgaagcgtc aactgctggc ggaagtacag tccatctgtc cggccggcgt gaccattatg 300
aacgtgcgtc agggcgaacc ttttaggttta ggccactcca ttttgtgtgc gcgacctgcc 360
attggtgaca acccatttgt cgtggtaactg ccagacgttg tgatcgacga tgccagcgcc 420
gaccgcgtac gttacaacct tgctgccatg attgcacgtt tcaacgaaac gggccgcagc 480
caggtgctgg caaaacgtat gccgggtgac ctctctgaat actccgtcat ccagactaaa 540
gagccgctgg accgtgaggg taaagttagc cgcatttttg aatttatcga aaaaccggat 600
cagccgcaga cgctggactc agacatcatg gccgttaggtc gctatgtgct ttctgccat 660
atttggccgg aactggaacg tactcagcct ggtgcattgg gacgtattca gctgactgtat 720
gctattgccg agctggcgaa aaaacaatcc gttgatgcaa tgctgatgac cggcgacagt 780
tacgactgcg gcaaaaaaat gggctatatg caggcggttg tgaagtatgg cctacgcaac 840
ctgaaagaag gggcgaagtt ccgtaaaggt attgagaagc tggtaagcga ataatgaaaa 900
tctgaccgga tgtaacggtt gataagaaaa ttataacggc agtggaaatt cgcagaaaa 960
gtaatttggtt gcgaatcttc ctggcggtgt tttatataaa ccatcagaat aacaacgagt 1020
tagcagtagg gttttattca aagtttcca ggattttcct tggccaga gcggattgg 1080
aagacaatta gcgttgaat tttcgggtt tagcgcgagt gggtaacgct cgtcacatca 1140

taggcatgca tgcagtgc tcggtagctgt aaagccaggg gcggtagcgt gcattaatac	1200
ctctattaa caaactgaga gccgcttatt tcacagcatg ctctgaagta atatggaata	1260
aattaagtga aaatacttgt tactggtgc gcaggattta ttggttcagc tgtagttcgt	1320
cacattataa ataatacgca ggatagtgtt gttaatgtcg ataaattaac gtacgccgga	1380
aaccggaaat cacttgctga tgtttctgat tctgaacgct atgttttga acatgcggat	1440
atttgcgatg cacctgcaat ggcacggatt ttgctcagc atcagccgga tgcagtgtatg	1500
cacctggctg ctgaaagcca tggtgaccgt tcaattacag gccctgcggc atttattgaa	1560
accaatattt ttggacttta tgccttttgaagccgctc gcaattactg gtctgcttt	1620
gatagcgaca agaaaaatag cttccgtttt catcatattt ctactgacga agtctatgg	1680
gatttgcctc atccagatga agtaaataat acagaagaat tacccttatt tactgagacg	1740
acagcttacg cgccaaggcag cccttattcc gcatccaaag catccagcga tcatttagtc	1800
cgcgcgtgaa aacgtacata tggttaccg acaattgtga ctaattgctc gaacaactat	1860
ggtccttatac atttcccggaa aaagcttatt ccactggta ttcttaatgc actggaaaggt	1920
aaggcattac ctatttatgg caaaggagat cagatccgatc actgggtgttga tggtgaagat	1980
catgcgcgtg cgtttatatac cgctgttaacc gaaggtaaag cgggtgaaac ttataacatt	2040
ggtgggcaca acgaaaagaa aaacatcgat gtagtgcctca ctattgtga ttggctggat	2100
gagattgtac cgaaagagaa atcttacgt gagcaaatac cttatgttgc tgatcgtccg	2160
ggacacgatc gcccgtatgc tattgtatgc gagaagattt gtcgcgcatt gggatggaaa	2220
ccacaggaaa cgtttgagag cgggattcgt aaaacggtgg aatggtacct gtcctataca	2280
aaatgggttg ataatgtgaa aagtggtgcc tatcaatcgt ggattgaaca gaactatgag	2340
ggccgccttgcgtt aatgaatatac ctccttttgc gcaaaacagg gcaggttaggt tgggaactac	2400
agcgtgctct ggcacccgttgcgtt ggttaatttgc ttgcctttgc tggactct actgatttt	2460
gcggtgattt tagtaatcctt gaaagggtgttag ctgaaaccgt aagaaggcatt cggccggata	2520
ttattgtcaa tgcagccgct cacaccgcag tagacaaagc agaatcagaa ccggagtttgc	2580
cacaattaat taacgcaaca agtgcgaaag cgattgcgaa agcagcaaataat gaagttggag	2640

cctgggttat ccattactcg actgattacg tcttccctgg aaatggcgat atgccatggc 2700
tggagacgga tgcaaccgca ccactaaatg tttacggtga aaccaagtta gccggagaaa 2760
aagcgttaca ggaatattgc gcgaagcatc ttatTTCCG gaccagctgg gtctatgcag 2820
gaaaaggaaa taacttcGCC aaaacgatgt tacgtctggc aaaagagcgt gaagaattag 2880
cggttattaa cgatcagttt ggtgcGCCaa caggtgctga actgctggct gattgtacag 2940
cacatGCCat tcgtgtcgca ctgaataaac cggatgtcgc aggcttgtac catttggtag 3000
ccagtggtac cacaacctgg tacgattatg ctgcgctgg ttttgaagag ggcgcAAAG 3060
caggcattcc cctgcactc aacaagctca acgcagtacc aacaacagcc tatcctacac 3120
cagctcgtcg tccacataac tctgcctta atacagaaaa atttcagcag aactttgcgc 3180
ttgtcttgcc tgactggcag gttggcgtga aacgaatgct caatgaatta tttacgacta 3240
cagcaattta atagttttg catcttgttc gtgatggtgg agcaagatga attaaaagga 3300
atgatgaaat gaaaatgcgt aaaggtatta ttttagcggg tggttctggt acacgtctt 3360
atcctgtgac tatggctgtc agtaaacagc tattacctat ttatgataaa ccgatgatct 3420
attacccgct ctctacactg atgttggcgg gtattcgcga tattttgatt atcagtagcac 3480
ctcaggatac tcctcgTTT caacaattgc tgggtgacgg tagccagtgg ggcctgaatc 3540
ttcagtacaa agtgcAACCT agcccagatg gcctcgcgca ggcatttatac atcggtaag 3600
agtttattgg tggtgatgat tgtgcttgg ttcttggta taatatctt tacggtcacg 3660
atctGCCaa gctaattggag gccgctgtta acaaagaaag tggtgcaacg gtattgcct 3720
atcacgttaa tgatccagaa cgctatggtg tcgttgagtt tgataaaaac ggtacggcaa 3780
tcagtctgga agaaaaaccg ttagaaccaa agagtaatta cggcgTTACA ggtctgtact 3840
tttatgataa cgacgtggtt cagatggcga aaaacttggaa gccgtctgca cgtggtgagt 3900
tagaaattac agatattaac cgtatTTATC ttgagcaggg acgtctgtct gtcgcgatga 3960
tggggcgtgg ctacgcgtgg ctggacacgg ggactcatca gagtctgata gaagcaagta 4020
atTTTATTGC gacaattgaa gagcGCCagg gattgaaggt ttccTGTcCT gaagagattg 4080
catttcgtaa aggtttatt gatgttgagc aagtaagaaa attagctgtta ccactaataa 4140

agaataatata tggcagttt ctttataaaa tgacgaaggaa ttcaaattaa tgaatgtgat 4200
tagaactgaa attgaagatg tgctaattct ggagccaaaga gtatttggtg atgatagagg 4260
tttctttat gagagctta atcaatcagc atttgaacat attctaggct atccggtcag 4320
cttggtaa gacaatcact cacgttcatc aaaaaatgtt ctcagaggcc ttcactttca 4380
acgcggcgag tacgcacaag ataaacttgt acgctgcact catggaggcag tttttgatgt 4440
tgctgttgat attcgaccca attcggtatc cttggtaaa tgggttggtg ttctgcttcc 4500
agctgataat aagcagcagt tgtggatacc aaaagggttt gctcatggct ttttggttct 4560
gtctgatatac gctgaatttc aatataaaac tacaaactat tatcatcctg aaagcgattt 4620
tggaaatatgt tggaaatgtt aacgcattgc aattgattgg ccccaaacat cagggtaat 4680
ccttcgcca aaagatgaaa ggctctttac gtttagatgag cttatcagat taaaattaat 4740
tgcattggc cggccttaag gaggactagt cccggcgcgc catgagtttta ataaaaaaca 4800
gtttttggaa ctttgcggg tatgtacttc cagctattgt gacactacca gctttggta 4860
ttatggggcg aaaatttaggc ccagaattat ttgggttattt cacttggca ttagctgttgc 4920
tgggttatgc aagcattttt gatgcaggcc ttactcgcc agtgatacga gaagtcgcaa 4980
ttgaaaaaga taatgaagaa aataagttga aaatttttc ttcaagcaca gttgttaatta 5040
tttatttgag ttggccgccc tcactcttattt ttttttttt tagtggtcat atcgatttgc 5100
tactgaacat tagtgagact ttttttcata atgttaagtgt ctcgcttaaa attctcgac 5160
catccataacc attatttttgc attactcaa tatggttgtc aatttttagaa ggtgaagaaa 5220
gatttggttt acttaatatc tacaaatcaa ttacgggagt gatatttagca atctcaccgg 5280
cattattttat acttattaaa ccctcttga tgtatgcgt aataggctt gttctagcaa 5340
ggttttatgt ttttatttttgc ttgttataaa ttgtcacga taaagtgcattt aaagctaaac 5400
taacaatcga tataccaaca attaaaaagat tgtttatgtt cggtgggttgg attacagttaa 5460
gtaatatcat cagccctgtt cttatcatatt ttgtatagggtt tattgtttca aatcaacttgc 5520
gggctgctaa tggtgctttt tatactgcac catcagaat tatttctcggtt ctttagtataa 5580
ttccagggtgc gttttcaaga gccttatttc caagattgc taatgcaat aattccgctg 5640

aaagatataa aacgaaaaga ttaattaca ttacacccat aataatcatc acccctat	5700
tttgtattgg cgtgttattt tcagagaaga taatgggttt atggatgggg gcatcattt	5760
ttggtagcc tggttggtt ttatcaatat tactgattgg ctttatTTT aatggattgg	5820
cacaagtacc atttgccagt attcaatccc gaggtcatgc taagataact gcatttgttc	5880
atctctttaga gttgttcct tatttattac ttttattttt cctcataaaaa gcacatgggg	5940
ttgttggcgc gggatttgcg tggcagtga ggatgatagt agattatata gcattaagtc	6000
ttttggacgg taagtatatt aataaataaa attcaaaatg caagtttata actcatggct	6060
ttatTTGGGT aggtgacaat ttataatgat atatataattt accttaactc ttcttctagt	6120
tatagccata atgtttctc ttctcgac aaaaaggtagg atcacatctc cattacctt	6180
gcattttta ccatggttac taacttaat tgcggata agtaattacg atcaattttt	6240
cgagtttaat gaaagaagct ttactctt gttgattgg ttacagttt ttttatattt	6300
ttatTCATA ggggaactgg ttaattataa acgtgaaaat ataaatgttt attatggct	6360
ttcacatatt aaatatgaat gtaaaaaata ttggatcatt gtcatccaa tttcattata	6420
taccattttc gaaatataata tgggtggat ggggggagca gatggattct ttctcaattt	6480
acgtcttgc aatacattgg agggctatac gggtaaaaaa tttatcttaa tgcctgctgt	6540
atatcctcta atgatggcta tggcgttcaat tggatgtctt acaaaaactt ccaaattaaa	6600
taaatactcc atttatttct ggtgtttt gtattgtatt ggcacaatgg gaaaattttc	6660
aatattaacg ccaatattga catatttaat tatttatgac ttcaaacata gattaaaagt	6720
aaaaaaaaaca ataaagttt cattgttcat aattatatta gcttaactt tgcattttac	6780
acgtatggct gagaatgacc actcaacatt tttatctatt ttagggctct atatttattc	6840
accaataatt gctttaggcc agttgaatga agtaaatagt agtcattttg gtgagttac	6900
gttttagattc atatatgcta taactaataa aattggcattt attaaagaat tgccagtaaa	6960
tactattctt gactattcat acgttccgtt accaacaat gtatatactg cacttcaacc	7020
attttaccag gatTTGGTT atactggcat catatttggc gcagtattt acggactaat	7080
atatgtgagt ttatacacgg ccgggtttcg tggaaataat acacaggcat tactgattt	7140

cgcattgttt tcagttagca gtgcaacggc tttcttcgct gaaacgctag taacgaattt	7200
agctggaaat gtgatgttag tattatgtac catcttacta tggcgattta cagtaatatg	7260
caaaccagta cagtaaccat tctaattggcc acctacaatg gcgaggcctt catcaaaaat	7320
cagatttgt cactacaaca acaaacattt tctaactggc gtttatttat tcaggatgat	7380
gggtctacag acaatactat atctataata aaaaacttcc aaaaatctga ctccagaatt	7440
cggtctatgg atgataattt gaaaggtcaa ggtgcagggaa aaaatttttt atcgctgata	7500
aagtacagcg agacagatta tacaatttat tgtgaccaag atgatatttg gttagaaaac	7560
aaaatatttg aatttagtaaa gtagcaaat gaaattaaat tgaatgtatc agatgcgcct	7620
tcgctatgtt atgctgatgg ctatgcttat atggatggtg agggtacaat cgattttct	7680
gggatatcta acaatcatgc ttagcaatta aaggattttc tttttttaa tggtgatcac	7740
caaggatgtt ctattatgtt caatcgatca atgaccaaattt ttcttctgaa ttatcgagga	7800
tttgtatatc tacatgacga tatcacaaca ttagctgcat acgctttgg taaagtttat	7860
tttctcccgaa aataccttat gttatataga cagcacacga atgcggtaac tggatcaaa	7920
acattccgca atggattgac ttctaaattt aaatcaccag taaactatct tttatcacga	7980
aaacattatc aggtaaaaaa atctttttt gaatgtaaca gctctatctt atcagagacg	8040
aataaaaaag tttttttgga ttttatttca ttttgtgaat caaataataa atttacagat	8100
tttttaagt tatggcgagg tgggtttaga ttaataaca gtagaactaa attattatta	8160
aaattcttaa tacggagaaa atttagcgaa tgatttcaat acttacacct actttatc	8220
ggcaacatac tttatcaagg ctattcaatt ctcttatatt acaaactgat aaagattttg	8280
agtggataat aattgatgat ggttgtatag atgcaacagc ggtacttgc gaagattttt	8340
gaaaaaaaaatg tgattttgac ttgattttt gctatcagga aaataatggt aagccatgg	8400
ctttaaacgc tgggtttaaa gcttgttagag gcgatttatat ctttattgtt gacagtgtat	8460
atgcactaac tcccgtatgcc ataaaattaa ttaaagaatc aatacatgat tgcttatctg	8520
agaaggaaag tttcagcgaa gtcggttta gaaaagcata tataaaaggg gggattattg	8580
gtaatgattt aaataattct tcagaacata tatactattt aaatgcgact gagattagca	8640

attnaataaaa tggtgatgtt gcatattgtt ttaaaaaaaga aagtttgta aaaaatccat	8700
tccccgtat agaagatgaa aaatttggtc cagaattata tatttggaaat aaaataactg	8760
acaaggcgaa gattcgattt aacataagca aagttatata tctttgttag tatcttgatg	8820
atggctttc taaaaatttc cataaccagc ttaaaaaaata cccaaagggg tttaagattt	8880
attacaaaga tcaaagaaaa cgagagaaaa cttatataaa aaaaacaaag atgctaatta	8940
gatatttgca atgttggat tatgagaaaa taaaatgaaa atactatttgc tcattacagg	9000
tttaggcctt ggaggtgctg agaagcaggt ttgtctttta gctgataaat taagtttaag	9060
cgggcaccat gtaaagatta tttcacttgg acatatgtct aataataaag tctttcctag	9120
cgaaaaataat gttaatgtca ttaatgtaaa tatgtcaaaa aacatttctg gagttataaa	9180
aggttgtgtc agaatttagag atgtttagc taatttcaaa ccagacattt tacacagtca	9240
tatgtttcat gcaaacatta tcactagatt gtctgttaatt ggaatcaaaa acagacctgg	9300
tattatatca actgcacata ataaaaatga aggtgggtat ttcagaatgc tcacatata	9360
aataaccgat tgTTTaaatgtt attgttgcattc aaatgttagc aaagaagcag tggatgagtt	9420
tttacggata aaagccttta atcccgctaa agcaattact atgtataatg ggatagatac	9480
caataaaattt aaatttgatt tattggcaag gagggaaattt cgagacggta ttaatataaa	9540
aaatgatgat atattattac ttgctgcagg tcgtttAACG tttagctaaatg attatcctaa	9600
tttattgaat gcaatgactc tgcttcctga acactttaaa cttattatta ttggatgtgg	9660
tgaattgcgt gacgaaatattt atatgcctt aaaaaatttgc caattatcta atagggtgtc	9720
cttggatggaa gttaaaaaaaaa atattgctcc ctatTTTCT gcatgtgata tttttgttct	9780
ctcttcgt tggtggatggat ttggattttt cgtggcagaa gctatgtcat gtgaggcgaat	9840
tgttgttggc acggatttcag ggggagtaag agaagtttattt ggtgacgatg attttcttgc	9900
acccatatct gattcaacac aacttgcag caaaattgaa aaattgtctt tgagccagat	9960
acgtgatcac attggttttc ggaatcgtga gcgtattttta aaaaatttctt caatagatac	10020
tattattatg cagtggcaag aactctatgg aactataatt tgctcaaaac atgaaaggtt	10080
gatttatattt tggAACGTGT ctTTTGTtTG aatttaatttcc aatctcaattt gagatTTTG	10140

tatttcaaaa ataccatcat agctaacgat gattggatt tatttaaga tgcttctat 10200
aaatatattg acgttttaa tgcgccaaa cgattggct gggAACAGAG aagtaaaact 10260
gttttgagaa tgaagagttt ttgagatgtt tatggatatt aaaaattgtat ccagtgaatt 10320
aattatTTT aataaatcaa gatttaatgt taataaatga taatctttc tgacactcat 10380
attaattatg agtggtagt ttggtaaacg gtAAACTATT atatgacAGC tagaacaact 10440
aaagTTTgc acttacaatt actcccactc ttaagtggcg ttCAAAGGGT AACATTAAC 10500
gaaatttagt cgTTtatatac tgattatgtat tatacactag tttgctcaaa AAAAGGTCCA 10560
ctaacaaaag cattgctgga atatgatgtc gattgtcatt gtatccccga acttacgaga 10620
gaaatttaccg taaagaatga ttttaaagca ttgttcaagc tttataagtt cataaaaaaa 10680
gaaaaatttg acattgtgca tacacattct tcaaaaacag gtatTTGGG gcgagttgct 10740
GCCAAATTAG CACGTGTTGG AAAGGTGATC CACACTGTAC ATGGTTTTTC TTTCCAGCC 10800
gcatctagta aaaaagttt ttacctttat ttttcatgg aatggatagc aaagtcttt 10860
acggataagt taatcgtctt gaatgttagat gatgaatata tagcaataaa caaattaaaa 10920
ttcaagcggg ataaagtttt tttaattcct aatggagtag acactgataa gttttctcct 10980
tttagaaaata aaatttata tagcaccttg aatctagtaa tggtagttttagt attatccaag 11040
caaaaagatc ctgagacatt attgcttgct gttggggggc tgctgaatga aaatgttaat 11100
gttaagctga cacttgttagg agatggtagaa ctaaaagaac agtttagaaag caggttcaaa 11160
cggaagatg gacgtataat ttttcatgga tggcagata acattgttaa tattttaaaa 11220
gttaatgtatc ttttatattt accttctttt tgggagggtt tgccattagc aatttttagaa 11280
gcattgagct gtggacttcc atgtatagtc actaatattc caggtataaa tagcttaata 11340
gaagatggct ataatggttt tttgtttgaa attagagatt gtcagttatt atctcaaaaa 11400
atcatgtcat atgttggtaa gccagaactg attgcacagc aatctaccaa tgcacgatca 11460
tttattctga aaaattatgg attagttaaa agaaataata aggtcagaca gctatatgt 11520
aattaagagc tcggtaaccg ggcctagggt gtaggctgga gctgcttcga agttccata 11580
ctttcttagag aataggaact tcggaatagg aactaaggag gatattcata tccgtcgacg 11640

gcggccgccc tgcaggcatg caagcttgcat ccatatggat cgcttagctta attaaataaa 11700
gccgttaagca tataaggcatg gataagctat ttatacttta ataagtactt tgtatacttta 11760
tttgcgaaca ttccaggccg cgagcattca gcgcggtgat cacacctgac aggagtatgt 11820
aatgtccaag caacagatcg gcgttagtcgg tatggcagtg atgggacgca accttgcgct 11880
caacatcgaa agccgtggtt ataccgtctc tattttcaac cgttcccgtg agaagacgga 11940
agaagtgatt gccgaaaatc caggcaagaa actggttcct tactatacgg tgaaagagtt 12000
tgtcgatatct ctggaaacgc ctcgtcgcat cctgttaatg gtgaaagcag gtgcaggcac 12060
ggatgctgct attgattccc tcaaaccata tctcgataaa ggagacatca tcattgatgg 12120
tggtaaacacc ttcttccagg acactattcg tcgtaatcgt gagcttcag cagagggctt 12180
taacttcatc ggtaccggtg tttctggcgg tgaagagggg gcgcgtgaaag gtccttctat 12240
tatgcctggt ggccagaaaag aagcctatga attggtagca ccgatcctga ccaaaatcgc 12300
cgccgtagct gaagacggtg aaccatgcgt tacctatatt ggtgccgatg gcgcaggtca 12360
ctatgtgaag atggttcaca acggatttga atacggcgat atgcagctga ttgctgaagc 12420
ctattctctg cttaaagggtg gcctgaacct caccaacgaa gaactggcgc agacctttac 12480
cgagtggaat aacggtgaac tgagcagtta cctgatcgac atcaccaaag atatcttac 12540
caaaaaagat gaagacggta actacctggt tgatgtgatc ctggatgaag cggctaacaa 12600
aggtacgggt aaatggacca gccagagcgc gctggatctc ggcgaaccgc tgtcgctgat 12660
taccgagtct gtgtttgcac gttatatctc ttctctgaaa gatcagcgtg ttgccgcattc 12720
taaagttctc tctggtccgc aagcacagcc agcagggcgc aaggctgagt tcatcgaaaa 12780
agttcgtcgt ggcgtgtatc tgggcaaaat cgtttcttac gcccaggcgt tctctcagct 12840
gcgtgctgctc tctgaagagt acaactggga tctgaactac ggcgaaatcg cgaagatttt 12900
ccgtgctggc tgcatcatcc gtgcgcagtt cctgcaaaaa atcaccgatg cttatgccga 12960
aaatccacag atcgctaacc tggcgtggc tccgtacttc aagcaaattt ccgatgacta 13020
ccagcaggcgc ctgcgtgatg tcgttgcttgc tgcagtacag aacggtattt cgggtccgac 13080
cttctccgca gcggttgcct attacgacag ctaccgtcgt gctgttctgc ctgcgaacct 13140

gatccaggca cagcgtgact attttggtgc gcataacttat aagcgtattt	ataaaagaagg	13200
tgtgttccat accgaatggc tggattaa		13228

<210> 18
<211> 13554
<212> DNA
<213> Artificial Sequence

<220>
<223> example 025B rfb locus nucleotide sequence - 025B-EPA production
strain stGVXN4459

<400> 18		
atgacgaatt taaaaggcagt tattcctgtta gcgggtctcg ggatgcataat gttgcctgcc		60
actaaggcga tacccaaaga gatgctacca atcgtcgaca agccaatgtat tcagtagcatt		120
gttgacgaga ttgtggctgc agggatcaaa gaaatcctcc tgtaactca cgcgtccaag		180
aacgcggctcg aaaaccactt cgacacctct tatgagttttag aatcactcct tgagcagcgc		240
gtgaagcgtc aactgctggc ggaagtacag tccatctgtc cggcgggcgt gaccattatg		300
aacgtgcgtc agggcgaacc ttttagttta ggccactcca ttttgtgtgc ggcacccgtcc		360
attggtgaca acccatttgt cgtggactg ccagacgtt tgatcgacga tgccagcgc		420
gaccgcgtac gttacaacct tgctgcatg attgcacgtt tcaacgaaac gggccgcagc		480
caggtgctgg caaaacgtat gccgggtgac ctctctgaat actccgtcat ccagactaaa		540
gagccgcgtgg accgtgaggg taaagtgcgc cgcattgtt aatttatcga aaaaccggat		600
cagccgcaga cgctggactc agacatcatg gccgttaggtc gctatgtgct ttctggcgat		660
atttggccgg aactggaacg tactcagcct ggtgcattgg gacgtattca gctgactgt		720
gctattgccc agctggcgaa aaaacaatcc gttgatgcaa tgctgatgac cggcgcacgt		780
tacgactgcg gcaaaaaat gggctatatg caggcgtttt tgaagtatgg cctacgcaac		840
ctgaaagaag gggcgaagtt ccgtaaaggt attgagaagc tggtaagcga ataatgaaaa		900
tctgaccgga tgtaacggtt gataagaaaa ttataacggc agtggaaatt cgcagcaaaa		960
gtaatttgcgtt gcgaatcttc ctggcggtt tttatataaa ccatcagaat aacaacgagt		1020
tagcagtagg gttttattca aagtttcca ggatttcct tggccaga gcggattgg		1080

aagacaatta gcgtttgaat ttttcgggtt tagcgcgagt gggtaacgct cgtcacatca 1140
taggcattgcata tgcatgcgtc tggttagctgt aaagccagggg gcggttagcgt gcattaatac 1200
ctctattaaat caaactgaga gccgcttatt tcacagcatg ctctgaagta atatggaata 1260
aattaagcta gcagtgaaga tacttggttac tggtggcgca ggatttattt gttctgctgt 1320
tgttcgtcac ataataaata atacgcaaga tagtggttt aatgtcgata aattaacata 1380
cgccggaaac ctggaatcac ttgcagatgt ttctgattct gaacgctatt tctttgaaca 1440
tgcggatatt tgtgatgcag ctgcaatggc acggattttt gctcagcatc agccggatgc 1500
agtgatgcac ctggcagctg aaagccatgt tgaccgttca attacaggcc ctgcggcatt 1560
tattgaaacc aatattgtgg gtacttatgt ccttttagaa gcccgtcgaa attattggtc 1620
tggtctggat gatgaaaaga aaaaaaaactt ccgtttcat catatttcta ctgatgaggt 1680
gtatggtgac ttacccatc cgatgaagt aaatagcaat gaaacgttgc cgctatttac 1740
ggaaacgaca gcatacgcgc caagtagtcc atattctgct tctaaagctt ccagcgatca 1800
tttggttcgc gcatggaaac gtacttatgg tttaccgacc attgtgacta attgctcgaa 1860
caactatggt ccttattcatt tccccggaaa gcttattcca ctggttattt ttaatttact 1920
ggaaggtaag gcattaccta tttatggcaa aggagatcag atcccgact ggttgtatgt 1980
agaggatcat gctcgagcgt tatataccgt cgtaaccgaa ggtaaagcgg gcgaaactta 2040
taacattgggt ggacacaacg aaaagaaaaa catcgacgta gtgttcaacta tttgtgattt 2100
gttggatgag atagtcccgaa aagagaaatc ttaccgagcgttcaacta atgttaccga 2160
tcgtccggga cacgatcgcc gttatgcgtat tgatgcgttcaacta atgttaccga 2220
atggaaacca cagggaaacgt ttgagagtgg gattcgtaaa acgggtggaaat ggtacctgtc 2280
caataaaaaa tgggttggata atgtgaaaag tgggtgcctat caatcggttcaacta 2340
ctatgagggc cgccagtaat gaatatcctc cttttggca aaacaggggca ggttaggttgg 2400
gaactacagc gtgctctggc acctctgggttcaacta atgttaccga 2460
gattactgtg gtgatTTTaaatcctgaa ggtgttagctg aaaccgttcaacta 2520
cctgatatta ttgtcaacgc agccgctcac accgcagtagt acaaaggcaga atcagaaccg 2580

aagttgcac aattactgaa cgcgacgagt gtcgaagcga tcgcgaaagc agccaatgaa 2640
gtcggcgct gggttattca ctactctact gactacgtat ttccgggac cggtaaata 2700
ccatggcagg aggaggatgc aaccgcacccg ctaaatgttt acggtaaaca caagtttagcg 2760
ggagaaaaag cattacaaga gcattgtgcg aagcaccta tttccggac cagctggtc 2820
tatgcaggt aaggaataa cttcgccaaa acaatgttgc gtctggcaaa agagcgtgaa 2880
gaattagccg ttattaatga tcagtttgt gcgc当地 actggctgat 2940
tgtacggcac atgctattcg tgtggcactg aataaaccgg aagtcgcagg cttgtaccat 3000
ctggtagcta gtggtagccac aacgtggcac gattatgctg cgctggttt tgaagaggcg 3060
cgcaaagcag gcattccct tgcaactaac aagctcaac cagtagccaac aacagcctat 3120
cctacaccag ctcgtcgtcc acataactct cgc当地ata cagaaaaatt tcagcagaac 3180
tttgc当地tgc当地tga ctggcaggaa ggc当地tgc当地ttaa cgaattattt 3240
acgactacag caatttaata gttttgc当地ttaa atggtaggagc aagatgtatt 3300
aaaaggaatg atgaaatgaa aacgc当地tgc当地ttaa ggtattattt tggc当地tgc当地tgg ttctggta 3360
cgtctt当地tac ctgtgacgat ggccgtc当地tgt aaacagctgt taccgattt tgataaaccg 3420
atgatctatt acccgctctc tacactgatg ttagc当地tgc当地ttaa ttgc当地tgc当地tata tctgattt当地tac 3480
agtacaccac aggatactcc tc当地tgc当地tcaaa caactgctgg gtgac当地tgc当地tgg ccagtgcc 3540
ctgaatctt当地tac agtacaaatg gcaaccgagc ccggatggtc ttgc当地tgc当地tggc gttt当地tattt当地tac 3600
ggtgaagagc ttattggc当地tgg tgatgattgt gctt当地tgc当地tac ttggc当地tgc当地tataa tatctt当地tac 3660
ggccacgacc tgccgaagtt aatggacgta gctgttaaca aagaaatgaa tgcaacggtta 3720
tttgc当地tctt当地tac acgttaatgaa tcctgacgt tatggc当地tgc当地tgg tggagtttgc当地ttaa taataacggt 3780
actgcaatta gc当地tgc当地tggaa aacccgctg gaacccaaaa gtaactatgc gttt当地tactgg 3840
cttt当地tattt当地tctt当地tac atgacaatgaa cgtt当地tgc当地tggaa atggc当地tggaa accttaagcc ttctgccc当地tca 3900
ggtgaactgg aaattaccgaa tattaaccgt attt当地tataatgg aacaaggacg tttgtctgctc 3960
gctatgatgg ggc当地tgc当地tggctc tgcatggctc gatacaggaa cgc当地tcaaaag tctt当地tattgaa 4020
gcaagcaact tcattgccac cattgaagag cgccaggaa taaaggttt当地tca ctgtccggaa 4080

gaaattgctt atcgtaaagg gtttattgat gctgaggcagg taaaagtatt agccgaaccg 4140
ttgaagaaaa atgcttatgg tcagtatctg ctc当地atga ttaaaaggta ttaataagat 4200
gaacgtaatt aaaactgaaa ttcctgatgt gctgattttt gaaccaaaag tttttgggga 4260
tgaacgtggc ttcttttg agagtttaa tcagaggatt tttgaagaag cagtaggtcg 4320
taagggtgag tttgttcagg ataaccattc taagtccagt aaaggtgtt tacgtggct 4380
tcattatcag tttagaacctt atgctcaagg aaaactggtg cgctgtgtt tgccgaggt 4440
ttttgatgtt gcgggtgata ttcgtaaattc gtcacctaca tttggaaat gggttggggt 4500
gaatttgcgt gctgagaata agcgtcagtt gtggattcct gagggatttgc cacatggttt 4560
tttgggtctg agtgatttag cagaagtttt atataaaacg aatcaatatt atgctccatc 4620
acatgaaaaa aatattatat ggaatgacct ctgcttaat attaaatggc cgagcacagc 4680
actgatcaact ctgtctgata aggatgcaaa tggggaaaga tttgaactaa gtgagtttg 4740
aaatgtctct cttaaaacat agtataatgga atgttgcggg ctactttata ccaacattaa 4800
ttgcaattcc cgcctttgga ttaattgcga ggaaaattgg ttagaacta tttgggttgc 4860
atacgtagc aatgattttt atagggtatg caagtatatt tgatgctggg ttaacaagag 4920
ctgttgtcg taaaaatgca ttactaaaaa acagagtggc cgattgtat acgataatag 4980
taacttctat tatcgctgtg atatttttag ggtttatcgg aggccccggg gtgtttctgc 5040
ttaaaggcga tattattgaa ctgttaataa tctcaccaat atattacgcc gattcgataa 5100
agtctcttagt attattatca tctctgatac ctgtattctt agtcacgcaa atactattag 5160
cagagcttga gggtcggaa tattttggaa ttctaaatatac aaaaaaaagt gtagggatt 5220
ctttaattgc agggttacct gcattatttg ttttaattaa tcaaacgctt tttctgcaa 5280
ttattgggt agcgattgca agagttatatactt gttgtgggtt aagctacatt atgagcagg 5340
aaagaataac tatcgatatac tcattttttt caataactgt tttaaagcgg ttatttagat 5400
atggcgggtg ggttaactata agtaacataa tatctcctat attagcgagt atggatagat 5460
ttattctatc ccatatccag ggagcatcaa aaatatcatt ctatacagtc cctaatgagc 5520
tggtaactag gcttggaaata gttccaggct ctcttggaa agctgtttt ccaaattaa 5580

gtcatgcaag gaattttaca gcgtcatatg cagagcaaaa aaaagcttat atattaatga 5640
ctgtcattgt aatgccttg gtttatttg tatattatta cgcaaagttt attttaacat 5700
tgtggatggg ggctgagttt gcagggattt cggtcgaaat attacggatt atgcttata 5760
ggtatatttt taactgttat tcacaaatct ctttgccaa catacaggcc tttggaaaag 5820
caaaatacac tgcatacatc catatgatgg aatttattcc ttatttgcata atgttatata 5880
taatttcaaa ggaatatggg gttattggtg ttgcgtggtt atggacaattt cgagtaataa 5940
ttgattttt gatgcttttataatgagtt atcggtgtaa taatcttataaaaaaagggt 6000
agcctgatga tatataattgtt ggtattaaat tggaatgggg ctatagatac cattaattgtt 6060
gttaaaagtt taatggattt aaatgttagc gattataaaaa ttatcattgt tgataactgtt 6120
tctatggata actcatatga tactataaaaa gaaaatctta attcattata tattgctgat 6180
aaaagtatca ttgaggtgaa gtatgaggat agaaataaat ataaaacctt agaaaacgat 6240
aaaatcatat taatacaatc tccgcaaaat aatgggtacg caagtggtaa taatattggc 6300
atagagttcg ctcttaatca ggagaatatg aaatacgtct gggttctgaa taatgatact 6360
gaagtggata aagaggctt aactcattta attagtaaat gtgattcaga taaaagtata 6420
gggatttgcg gttctcggtt agtctatttt gccgacagag agatgcagca aggacttaggt 6480
gggggtgcata acaaatggtt atgcactaca aaaaattatg aaatgggaag attagttcc 6540
aaaaaaatatg atgatgaagt cattagtaat gatatagatt atataattgg cgcatcgatg 6600
tttttctcta gagaatggtt ggaaacagtt ggattgatga atgaagaata ttttttatac 6660
tatgaagagt tagatatttg cctcagagca aaagcaaaga actttaaattt aggtatttgc 6720
tcagaaaagtt tggtttatca taaaataggt gcaagtactg atgggggaaa gagcatgatg 6780
gctgatctt gctcaataaa aaataggctg gtcattacag aaaggttta tccccaaat 6840
tattggacgg tatggttgtc actttttgtt gtagcattta accgtgctag aagaggtgag 6900
tttaataaga tgaaaagatg tttgaatggtt atgtttaact tcaaacgaaa caaaggtagc 6960
aaatgccatt agaatatgca cttaatcatg gtgttaataa atctatagtt tgatatgtta 7020
ttaaagggtt ttaatgaaa gtggctttt tatctgctta tgatccacta tctacatcca 7080

gttggctgg cacacccat tatatgctaa aggcatatc gaagagaaaat attccattg 7140
aaatattagg accggtaat agctatatga tatacatgtt aaaagtataat aaattaat 7200
taagggttt cgaaaaagaa tatgattata gtcattcgaa gttgcttcc aggtattacg 7260
tagaaatatt cggttagaaaa ttaaaaaaaaa ttgatggttt ggattttatt atcgcacctg 7320
caggttcctc acaaattgct ttttaaaaaa caaccatacc aataatataat ctatcgata 7380
caacatatga tcaattaaaa agctattatc cgaatttaaa taaaaaaaaaca attataaatg 7440
atgaggatgc aagtttaatc gaacgcaagg ctattaaaa agcaacagta gtatcttcc 7500
catctaaatg ggcaatggat ttttgcagga attattacag attagatttt gataaattag 7560
ttgaaatacc atggggggct aatttatttg atgatattca ctttgctaatt aaaaatataa 7620
ttcaaaagaa tagttatact tgtctttct tgggagttga ttgggaaaga aaagggtggga 7680
aaacagcctt gaaagcaatt gaatatgtaa ggcagttata tgggatcgat gttagactaa 7740
aaatttgtgg atgtactccg aatcaaaga ttttacctac ttgggttgaa ttaattgata 7800
aagtagataa aaataacgtt gacgaatatc agaaattcat cgatgtgtta tctaacgctg 7860
atatacttct tttaccaacc attgctgaat gttatggaat ggtattttgt gaagctgctg 7920
ctttggatt gcctgttgc gctacagata caggtggagt cagttctata gttatcaacg 7980
aaaggacggg gatattaatt aaagacccgt tagactataa gcactttgga aatgcaattc 8040
ataaaaataat tagttccgt a gagacttac aaaactactc cccaaacgca agaatttagat 8100
ataataatattt attgcattgg gacaattggg cttaaaagat aattgagatt atgtatgagc 8160
ataagaatag aagaatcaa tagcacaaaa agaattataat gtttattttt actttttctt 8220
gtttccctg atttttgtt ttatacatta ggggttgata atttttagcat ttcaacgata 8280
atctcaatta cattgcttt tgttttta agagctaaaa atattgcaa agataatttt 8340
ctaataatag tagcgttatt catattgtt tgtttaact gtttgttaag tatgctattt 8400
aatattgaac aggcttaac atttaaagtt gtacttcaa tatatagcat cttaataatg 8460
gcatacgctc cctcttgtt tgcacagacg ttgtggttat gttctgaaga aatacttaag 8520
agatccgtct tttatggtt cgcatctt tgccttatttgcattataag tattctttta 8580

cagaagactg agattataca tgataaaagt atgattctt ttcctgaacc atcagcattt 8640
gcattggttt ttatacctat ctttcattt tgtttatact atacaagagg gggggggcta 8700
ctattgctct atatattatc tttgggtatt gcgttaggta tccagaattt aacaatgttg 8760
gtaggcattg tgattagtg ttttgtgatg aaaaaaataa ctataaggca aactattgtt 8820
atactttgg gggcatggat ttttccatg atattaagtg atttagacat ttcttactat 8880
acatcgccgc ttgattttaa aaatactacg aacctatcag tgcttgata tccttcagga 8940
attgaaagag ctttcttcaa ttttattaca agttatggtc ttggtattgg tttcaacaa 9000
atgggagtga atggggagat aggaatatat caacaattt tagctgaact tgatccccct 9060
atgttaaata tatacgatgg ctcatttatt tcctctaagt taatatctga gtttggggtt 9120
attggtgcat taatgtgtat tttctatTTT ttttattttt cccgattttt tctgcgtttc 9180
aaaaaaaaagta agagatattc accgcagtat attttagcat atagcttcta catgtgtttc 9240
ttcatccctc ttttatacg tggtgctggt tatataaacc cctatgtgtt tatgttattt 9300
tcatcaatat ttttgtgcaa atatcacgct aaaaatatct tgatgaaatc taatgtccag 9360
atagctatat aatagtagat tatattatca ttatcacgta aattacatat taatagcata 9420
tatgataact aggacataaa taatgtgcat taaaaaaaaa cttaagttaa ttaaacgata 9480
tggccttat ggtggtctta ggcttcttaa agatatattc ttaacaaaat ttttattttg 9540
ttcaaatgtt aggattatta gatttccatg ttatattaga aaagatggaa gtgttagttt 9600
tggaaaaggt tttacatcag gtgttaggatt acgagttgat gcatttatgg atgccgtatg 9660
ttccattgga gaaaatgttc aaattaatga ctatgttcac atcgccgcta ttaataatgt 9720
cattattggt agagatacat taatagcaag taaagtattt attagtgatc ataatcatgg 9780
tatttttctt aaatccgata tccatagttc accaactatt attccttcgt ctagccccct 9840
tgaatctgca cctgtgtata ttggagagcg tgtgtggatt ggcgaaaatg tgacaatatt 9900
accaggtgcg tgtataggtt atggtaggtt tattggcgca aacagtgttg ttcgtggta 9960
gattcctaattt aatgtgatca ttgctgggtt tccagctaaa attgttaaaa aatataacta 10020
tgagcgtatg caatggaaaa gaatatagtt gtaatatcgg ctgttaattt tacaaccgga 10080

ggcccctta ccgtactaaa aaatgtgctt acagcaacta aagatagagc cgaatgtaaa 10140
tttattgcac tggttcatag ctctgctgaa ctaatggaat tatttccgtg gggtgaattt 10200
atagagtatc cagaagtcaa gtcttcgtgg gttaaaagat tatatttcga atatataact 10260
tgcaatagat tatctaaggt gattaaggca actcattggg tatgcttaca tgatattaca 10320
gcaaatgtta gtgtacccta tagattgtt tattgccaca atcctgcacc gttctataaa 10380
tatttaagct atcgagatat tataggagaa cctaaatttt atcttttta tccttttat 10440
gggctttat acaatatcaa tataaaaaag aacacagcag ttttggcca gcagcagtgg 10500
ctaaaaaaaaaag aattcgaaaa aaaatataag ttaaagaatg ttgttggtag tcgcccgtaa 10560
gatatttgcc ctttgaaag tcatggtttgc gtaagaaata ataataaaaa ggatgtgagg 10620
atattttacc cagcagtgcc ccgtatattt aaaaactttg aagttatcat acgtgctgca 10680
caaataattac aagataaaaaa tattcatttt tatcttactt ttgatggtag tggaaataag 10740
tatgcaaaaa gaatataaa attagcttcc gaactgaaaa atgtacattt cctcggttac 10800
cttaatgcaa ccgagatggta taactttat caagattcag atatttttg tttccatcg 10860
aaactagaaaa cgtggggatt accattatca gaagctaaaa catacaaaaa atggatattt 10920
gcggcagact taccttatgc tcatgaagtt ttatataact attcaaaaac tagatatttt 10980
ccatttgacg atgagaaaaat acttggtcgc tacatattag agtacacaag taaaatatg 11040
catgaagata taaaaatag tagggtaat ttataataatg atgcattgac tggtttgaa 11100
cagtttattt aatatatcct caaggggaac tgacgtgggt tatattataa tcgtttcaca 11160
tggccatgat gactatatac aaaaatttt attaaattt aagttggccct ctggaaagatt 11220
taaaataata gttcgtgata acaaaaatgtc aatggtttta aaaaaaacat gcgaaaaaaa 11280
ttgcgttaacc tatttgcattt gagggcaata tggatttgaa cataataata acatagcagt 11340
gtcatatata attaataact tcatgatttata gaataatgtat tattttctct ttcttaaccc 11400
cgatgtattc ataaccagtgc aaagtttgcattt taattatgtt gattatataa ttagtaatga 11460
ttataagttt agcacattat gtctttatgc agatttact aaaaagcaac atgattattc 11520
aatacggagt tttccaaactt tataatgtt tctttgttct ttatttattgg gggtgaataaa 11580

aagtaaaatt aagaaggaaa atatactttc tgatactgta gttgattgg gtgctggctc 11640
atttatgctt attcatgctt taagttctt aaatgtaat ggaaaaatgtc aaaaatattt 11700
tatgtattgt gaagatattg acctttgtat gcgtttaaaa ttaagtggag tagatctta 11760
ctatactccc cattttgatg ctattcatta tgcgacat gaaaatagaa gaatatttac 11820
taaaggcattt cgatggcata taaggagtat tacgcgtac atattacgga aaccaattct 11880
ttcttataaa aactatagaa aaattacatc cgaactggta aagtgattaa ggatccgtgt 11940
aggctggagc tgcttcgaag ttcctatact ttcttagagaa taggaacttc ggaataggaa 12000
ctaaggagga tattcatatg gataaagccg taagcatata agcatggata agctatttat 12060
actttaataa gtactttgta tacttatttg cgaacattcc aggccgcgag cattcagcgc 12120
ggtgatcaca cctgacagga gtatgtaatg tccaagcaac agatcggcgt agtcggatg 12180
gcagtgatgg gacgcaacct tgcgctcaac atcgaaagcc gtggttatac cgtctctatt 12240
ttcaaccgtt cccgtgagaa gacggaagaa gtgattgccg aaaaatccagg caagaaactg 12300
ttcccttact atacggtgaa agagttgtc gaatctctgg aaacgcctcg tcgcattctg 12360
ttaatggta aagcaggtgc aggacacggat gctgctattt attccctcaa accatatctc 12420
gataaaggag acatcatcat tcatggtggt aacacccct tccaggacac tattcgtcgt 12480
aatcgtgagc tttcagcaga gggcttaac ttcatcggtt ccgggtttc tggcggtgaa 12540
gagggggcgc taaaagggtcc ttcttattatg cctggggcc agaaagaagc ctatgaattt 12600
ttagcaccga tcctgaccaa aatcgccgcc gtagctgaag acggtaacc atgcgttacc 12660
tatattggtg ccgatggcgc aggtcaactat gtgaagatgg ttccacaacgg tattgaatac 12720
ggcgatatgc agctgattgc tgaaggctat tctctgctt aaggtggcct gaacccacc 12780
aacgaagaac tggcgcagac ctttaccgag tggaaaacg gtgaactgag cagttacctg 12840
atcgacatca ccaaagatat cttcaccaaa aaagatgaag acggtaacta cctgggtgat 12900
gtgatcctgg atgaagcggc taacaaaggt accggtaaat ggaccagcca gagcgcgtg 12960
gatctcggcg aaccgctgtc gctgattacc gagtctgtgt ttgcacgtt tatctttct 13020
ctgaaagatc agcgtgttgc cgcatctaaa gttctctctg gtccgcaagc acagccagca 13080

ggcgacaagg ctgagttcat cgaaaaagtt cgtcgtgcgc tgtatctggg caaaatcggtt	13140
tcttacgccc agggcttctc tcagctgcgt gctgcgtctg aagagtacaa ctgggatctg	13200
aactacggcg aaatcgcaa gatttccgt gctggctgca tcatccgtgc gcagttcctg	13260
cagaaaatca ccgatgctta tgccgaaaat ccacagatcg ctaacctgtt gctggctccg	13320
tacttcaagc aaattgccga tgactaccag caggcgtgc gtgatgtcgt tgcttatgca	13380
gtacagaacg gtattccggt tccgaccttc tccgcagcgg ttgcctatta cgacagctac	13440
cgtgctgctg ttctgcctgc gaacctgatc caggcacagc gtgactattt tggtgcgcat	13500
acttataagc gtattgataa agaagggtgtt ttccataccg aatggctgga ttaa	13554

<210> 19
<211> 15197
<212> DNA
<213> Artificial Sequence

<220>
<223> example 075 rfb locus nucleotide sequence - 075-EPA production strain stLMTB11737

<400> 19
atgacgaatt taaaaggcagt tattcctgta gcgggtctcg ggatgcataat gttgcctgcc 60
actaaggcga tacccaaaga gatgctacca atcgtcgaca agccaatgtat tcagtagatt 120
gttgacgaga ttgtggctgc agggatcaaa gaaatcctcc tggtaactca cgcgtccaag 180
aacgcggctcg aaaaccactt cgacacccctt tatgagtttag aatcactcct tgagcagcgc 240
gtgaagcgtc aactgctggc ggaagtacag tccatctgtc cggccggcgt gaccattatg 300
aacgtgcgtc agggcgaacc ttttaggtta ggccactcca ttttgtgtgc gcgacccgtcc 360
attggtgaca acccattttgt cgtggtaactg ccagacgttg tgatcgacga tgccagcgc 420
gaccggctac gttacaacct tgctgccatg attgcacgtt tcaacgaaac gggccgcagc 480
caggtgctgg caaaacgtat gccgggtgac ctctctgaat actccgtcat ccagactaaa 540
gagccgctgg accgtgaggg taaagtcaac cgcattgttg aatttatcga aaaaccggat 600
cagccgcaga cgctggactc agacatcatg gccgttaggtc gctatgtcgt ttctgcctgat 660

atttggccgg aactggaacg tactcagcct ggtgcattgg gacgtattca gctgactgat 720
gctattgccg agctggcgaa aaaacaatcc gttgatgcaa tgctgatgac cggcgacagt 780
tacgactgcg gcaaaaaaat gggctatatg caggcgttt gtaagtatgg cctacgcaac 840
ctgaaagaag gggcgaagtt ccgtaaaggt attgagaagc tgttaagcga ataatgaaaa 900
tctgaccgga tctaaccggtt gataagaaaa ttataacggc agtggaaaatt cgccagcaaa 960
gtaatttgtt gcgaatcttc ctgccgttgt tttatataaa ccatcagaat aacaacgagt 1020
tagcagtagg gtttattca aagtttcca ggattttcct tgttccaga gcggattgg 1080
aagacaatta gcgttgaat tttcgggtt tagcgcgagt gggtaacgct cgtcacatca 1140
taggcattgca tgcagtgc tcggtagctgt aaagccaggg gcggtagcgt gcattaatac 1200
ctctattaaat caaactgaga gccgcttatt tcacagcatg ctctgaagta atatgaaata 1260
aattaagcta gcagtgaaga tacttgttac tggtgccca ggatttattt gttctgcgt 1320
tgttcgtcac ataataaata atacgcaaga tagtgttgg aatgtcgata aattaacata 1380
cgccggaaac ctggaatcgc tcgctgaaat ttctgattct gaacgttatt catttgagca 1440
tgccagatatac tgcgtgcgc aagcgatggc tcgtatccc gcacagcacc agccagacgc 1500
ggtgatgcac ctggcagcag agagccacgt tgaccgctca ataactggcc ctgcccatt 1560
tattgaaacc aatattgtgg gtacttatgt tcttttagaa gcggcgccca attattggtc 1620
tggtctggat gatgaaaaga aaaaaaactt ccgcattcat catatttcta ctgatgaggt 1680
gtatggtgac ttacccatc cggatgaagt aaatagcaat gaaacgttgc cgctatttac 1740
ggaaatgaca gcatacgcgc caagtagtcc atattctgt tctaaagctt ccagcgatca 1800
tttggttcgc gcatggaaac gtacttatgg tttaccgacc attgtgacta attgctcgaa 1860
caactatggt ctttatcatt tccccggaaaa gcttattcca ctggatttcc ttaatgcact 1920
ggaaaggtaag gcattaccta tttatggcaa aggagatcag atccgcgact ggttgtatgt 1980
agaggatcat gctcgagcgt tatataccgt cgtaaccgaa ggtaaagcgg gcgaaactta 2040
taacattggt ggacacaacg aaaagaaaaa catcgacgta gtgttcaacta tttgtgattt 2100
gttggatgag atagtcccga aagagaatc ttatcgtgag caaattacct atgttgctga 2160

tcgcccaggg catgatgcc gttatgcaat tgatgccat aaaatttagcc gcgaattggg	2220
ctggaaacca cagggaaacgt ttgagagcgg gattcgtaaa actgtggaat ggtatctgtc	2280
caatacaaaa tgggttgata atgtgaaaag tggtgccat caatcggttga ttgaacagaa	2340
ctatgggggc cgccactaat gaatatcctc cttttggca aaacagggca gggtgggttgg	2400
gaactacagc gtgctctggc acctctgggt aatttgatttgc ctcttgatgt tcactccact	2460
gattactgttgc gtgatttttag taaccctgaa ggtgtggctg aaaccgttag aagcattcgg	2520
cctgatatta ttgtcaacgc agccgctcac accgcagtag acaaaggcaga atcagaaccg	2580
gagtttgcac aattactgaa cgcgacgagt gtcgaagcga tcgcgaaagc agccaatgaa	2640
gtcggcgctt gggtttattca ctactctact gactacgtat ttccggggac cggtaaaata	2700
ccatggcagg aggaggatgc aaccgcaccg ctaaatgttt acggtaaaac caagtttagca	2760
ggagaaaaag cattacaaga gcattgtgca aagcaccta tttccggac cagctgggtc	2820
tatgcaggta aaggaaataa cttcgccaaa acgatgttgc gtctggcaaa agagcgtgaa	2880
gaattagccg ttattaaatga tcagtttggc gcgccaaactg ggcgcaggtt gctggctgat	2940
tgtacggcac atgccattcg tgtggactg aataaaccgg aagtcgcagg tttgtaccat	3000
ctggtagcca gtggtagccac aacctggcac gattatgctg cgctgggttt tgaagaggcg	3060
cgcggccat gcattccct tgcactcaac aagctcaacg cagttaccaac aacagtctat	3120
cctacaccag ctcgtcgcc acataactct cgccttaata cagaaaaatt tcagcagaac	3180
tttgcgcttgc tcttgcctga ctggcagggtt ggtgtgaaac gcatgctcaa cgaattattt	3240
acgactacag caatttaata gttttgcattt cttgttcgtt atgggtggaaac aagatgaatt	3300
aaaaggaatg atgaaatgaa tacgcgtaaa ggtattattt tagcgggtgg ttctggta	3360
cgtctttatc ctgtgactat ggctgtcagt aaacagctgt taccgattta tgataaaccg	3420
atgatctatt acccgctctc tacactgatg ttggcgggtt ttcgcgatat tttgattatc	3480
agcacgcccac aggataactcc tcgtttcaa caactgctgg gtgatgggag ccagtgcccc	3540
ctaaatcttc actacaaagt gcaaccgagt ccggatggtc ttgcgcaggc atttatcatc	3600
ggtaagagt ttatcggtgg tcatgttgcgtt gctttggta ttggtaaa tatcttctac	3660

ggtcacgacc tgcctaagtt aatggatgcc gctgttaaca aagaaagtgg tgcaacggta 3720
tttgcctatc acgttaatga tcctgaacgc tatggtgtcg ttgagtttga taaaaacggt 3780
actgcaatca gcctggaaga aaaaccgtta caaccaaaaa gtaattatgc ggtaaccggg 3840
ctttatttct atgataacta cgttgtggaa atggcgaaaa atcttaagcc ttctgcccgc 3900
ggtgaactgg aaattaccga tattaaccgt atctatatgg aacaggggca tttatctgtt 3960
gccatgatgg gacgtggata tgcctggctg gacacgggga cacatcaaag tcttattgaa 4020
gcaagcaact tcattgccac cattgaagag cgccaggcgt taaaagtttc ctgccccgaa 4080
gaaattgctt accgtaaagg gtttattgat gctgaggcagg taaaagtatt agctaaaccg 4140
ctgaaaaaaaaa atgcttatgg tcagtatctg ctaaaaatga ttaaaggtta ttaataaaaat 4200
aatgttatt aaaacagaaaa ttccagatgt actgattttt gaaccgaaag tttttggta 4260
tgagcgtggt ttctttatgg aaagctttaa tcagaaagtt ttcaagagg ctgttagggcg 4320
gaaggttgaa ttgttcagg ataatcattc taaatcgtgt aaagggtgtac ttagaggttt 4380
acactttcag cttccccc ttgagcaggc aaaattagta aggtgtatag ttggcgaggt 4440
atttgatgtt gcagtagaca ttagacctaa ttctgaaaca ttgggtcat gggttggagt 4500
aactcttcg tcagaaaata aaaggcagct atggattcca gaaggattcg cccatggttt 4560
tttaacttta agtgatattg cagagttgt ttataaaaact aacaactatt attctttaaa 4620
tcatgaaagg ggagtcattt ggaacgatga ggaaattaac attgcctggc cctctcaatc 4680
agagaagatt ctgtcacaga aagatattaa tttaccatca ttttagatttgc ttcaaatttt 4740
tagcaagtag tgttatctt acactgcaca tagtcatcat tttttatgct ttaagtaaat 4800
tatattgcac atctataaca caaagcgcaa taatattcg acctgatgaa ggtttgggt 4860
tatttatctt tctaggcggtt ttttatgact aaaatagttg tggtttctac agctccaata 4920
ttcccgacaa ataatggta caaaagttct gtatttagaa gaattgatga gttattaaat 4980
gaggataatg aggtcggttt gattgaaata aaccttgaaa atgttacgga aaagaaagat 5040
gaattaatac caacaagatt taataatatt caaagatatg aagtaaaaaa aatatctaga 5100
tcatttattt ccgagttaca aatatttattt gatatcagaa ctcggatgtaca acaattttt 5160

tcttctgctg acattagaga taacataaaa aagataattg atttagaaaa accttctatt	5220
attattgctg agtctatatg ggcgttgcaa gcattgccta ttgaaattag tgcgagaata	5280
cactgtgtta ttcatgatgt ggcaactgat ttctttaaag aaatgttgt atctcataat	5340
gaggttgtac gaaaaatttt gtttttaat gattacctaa agttgaaaat tactgaagaa	5400
aatattatca aacgttgag agttgagcaa tttatcttc tgacagaaga agataaatgt	5460
tggtataaaa caagatacaa tattgatgag gttgttggt ccttagcgag caatcatctt	5520
tatgttagaaa agattaagag aactatcaat ttccaaaccc cttcctgct tattccggc	5580
agcattgaat ttccacaaaa ttttacggc ttaatttgtt ttataaaaaa tatatacct	5640
ggattaaata gaaaaataag aatagttgta acaggaaagg catcagataa aaaaataaaag	5700
atgttaaact gtggagagga aattacctt acgggagagc ttgactttc cacatataat	5760
aaacttagct caacatgctt gtgtgttatt gcaccgatta caacggcac tggattaaa	5820
ataaaaaat tagaagctgt aaaaaaggt attcctgtac ttacaacaaa atttgcttca	5880
aaaggaatat gttccgattt atgttttat tgcgaggagg atactgacac aaactttgtc	5940
aatttaatta acagtttct tgaacgaca ttaagagtcc aagaatgaat ttattgcttt	6000
tttcagtcct tgcgttggt ttaatattgg ctggccca taataataaa agtggagata	6060
ttaacgcata cttaatgttt ttctcgtgg tcctaatttgtt attaatatca gggctgcgt	6120
tgaatgatag tgattatatc gaatacagga aaatgtataa tgaagtgcct attttatgt	6180
actttagtct cgcatctata agagatatac atggggaggt aggctatcta ttcttatcat	6240
caatcttaa aactttatgc ttgccatttc aattatttct ttttttatt gctttttat	6300
cactcctgct tacatatttt tcattcagaa aaataagttt aataccgata ctatcgtag	6360
tttttattt aagccatgct tttatagtta gagatttgat tcaaatttagg gcaggattag	6420
ctgttagcat atcattatat tcaataatta aatttaaagg aaataaaagt ataattacag	6480
gagttttatt tgcttcttg attcattctg gggcgcttattattgttctt tgttatcctt	6540
ttttcaaaaa aaaatacata acattaaaaa tgatgttggtt tttatTTTA gtgtcaatta	6600
ttttttctta tttgaatggg cttaaatttat cgataacaact cttatctcaa tatagttgc	6660

ttccaactgc aatttcgaat tatgttggtt gggagaata tgattatcgg gtgagttat	6720
ttactaatcc ggaaaaattt aaagggtgtt ttttaattgt cttaatgcac aaatatgtac	6780
tttcagatat taaaatgag aaaattatag tgcttataa ctatatgtt ttaggtgtat	6840
tagctatggt tgcattgagt gggatggcta ttcttcagg ccgtcttca tccttctga	6900
cactaggtga aagcattta attgtatatg ctctgttcta caaaagaaat acacctctgg	6960
cgtttcta atttttttt ttaacaattt tgcaatttgg atatgtatcta tttatttcta	7020
atgtgcattcc tgagcttact ctgattatat ttgggtgaat ctaagtgaaa aataataaaa	7080
taggcatact tatctctaaa atacaaaatc ttggacctgt gaatgttagt cgaggattga	7140
taaaagaaaa taaaaatat gctttactg tttttgttt aacaaatagc gtagataaaa	7200
atatatatga tgagttatgc tgtaggttgg ccaaggttat attaatacca gatggactt	7260
ggttcagcaa aatttttattt gtgagaagtt ttttaaagga acatccacat aatatcttac	7320
attcacatgg gatcacggcc gatatgtttt cttacttct gaatggcgtg aaaatatcta	7380
ctattcacaa tagacttagt gaggattata tcccattatt tggcgcggtt aaagggatg	7440
ctatataattt tcttcatcgt tttatattac gaagattaa tcatatcggt gcttgctcag	7500
cagcggtcca atcaaaactg aaacaatcga aagtaaaaac taaaataacc accatccaga	7560
atgggattga tataacttagg tttaagacac ttgagttcga taaaaaaaaa ttatttgggg	7620
aaaaacacgg atttgatagt gaaaaaagaa tatttatata ttgtggctcg ttatcattaa	7680
ggaaaaatat tgcttaccc ttggAACACT tagccatcga agaaaatgtat atattttaa	7740
ttcttaggtga tggtaactt ttttagatatt gtaaggataa atattctaaa gatttacggt	7800
atatatattt gggaaagtt gaatgccctc ttgaatatta tcaattatca gatatttttg	7860
tttccgcttc tttatcgaa gggctccct tggcactatt agaagctgcc tctactgggt	7920
gctatttata tgtagcgat atagagcccc atagagaat tgcatctcta ttaggagagg	7980
aaaatatttc tatgtttaaa attaaggatg gatcatataa ttatggcaa cctaaaataa	8040
aaaaagctga ctataacgct ctttctgacg ataaacttta caatataatcc gataaaaaaa	8100
tgtcaaatct ttatgacaaa cttttgttt ctttattaga gcagaggcac taatataatg	8160

atttatgttt cggttaattc tcatggtcat ttcaaaaactc ttaaggaatt aggaggaga	8220
tcaaaaattaa ataatcacag cagaattaaa gttatcatca aagataattt aggagagagc	8280
gagctttgg attttgtca ggaaaacaaa ataacttatt taaggtctaa agagaaaaaa	8340
ggatttggag agaataataa tgaagtttt tcctctatat cctccttaat tactaaggaa	8400
gattttttg tggttatgaa tcctgatata tatattgagt gctctgatct attagatgtc	8460
gtagatgagt gtggttcagc gaatgttaat ctagcaacga taaatttata cagggatttt	8520
gataaaaaaa catatgataa ctcagtaagg aaattccct cggcaattga ttttttatg	8580
tcattttat ttaagaaaaa tgactgtgta gtaaataaga acaaataac gaaaccaaca	8640
tatgttGatt gggctgcagg ttctttcta atatttaatg cttctttta ttcaaaaactc	8700
aacggattca acgaaaagta ttttatgtat tgcaagata ttgatataatg ttggcgagct	8760
aaaaaacact tcaatacttc agtttatac tatccatgct atgcagcaat tcatttggca	8820
caatttaaca atcgttaggat ttttagtaga catttcattt ggcataaaaa aagtattatc	8880
cttttttat tatataaaaaa tggtagtgctg cgttctagta agttgcttta atgctaataat	8940
tcttttaaga ggtgagaatg atacctgtta ttttggtgg tggttcggga agtcgcttgt	9000
ggccactttc acgagaaaaag ttccccaaagc agttttaaa gttgactggc agtttgacaa	9060
tgttgcagtc aacattgtca cgtcttaata atttaaatgc ttagtattca atagttatata	9120
gcaacgaaga gcatagattt attgttgcag aacaattaag agagtttaggc aaactttcaa	9180
ataacattat tcttgaaccc aaaggtcgta atacagcccc tgctataaca ctgcagcat	9240
tagcagcaaa aagaaaattc gctgatgaag atccattgat tcttattttt gctgcagatc	9300
acaacatcca agacgaacat gtttctgtg aggcaattaa taaggcgtca tcttagcta	9360
gttatggaaa actagtgact tttggtatcg ttccattcaa acctgaaact gggtagggct	9420
atattcgtcg cggtagtggaa gtgcctgttag atgagcagca tgccgtggcc tttgaagtgg	9480
cgcagttgt cggaaaaccg aatctggaaa ccgcgcaggc ctatgtggca agcggcgaat	9540
attactggaa cagcggatgt ttcctgttcc gtgccggacg ctatctcgaa gaactgaaaa	9600
agtatcgccc ggatattctc gatgcctgtg aaaaaggcgat gagcgccgtc gatccggatc	9660

tcgatttat tcgtgtggat gaagaggcgt ttctcgcttg tccggaagag tcgggtggatt 9720
acgcggtcat ggaatgcacg gcagatgccg ttgtggtgcc gatggatgcg ggctggagcg 9780
atgtcggttc ctggtcttca ttatggaga tcagcGCCa caccGCCGAG ggcaacgttt 9840
gccacggcga tgtgattaat cacaAAactg aaaacagcta tgtgtacGCC gaatctggcc 9900
tggtcaccac cgTCGGGgtg aaagatttgg tggttagtgca gaccaaaagat gcagtgcgtga 9960
ttGCCGACCG taatgcggtg caggatgtga agaaagtggt cgagcagatc aaagctgatg 10020
gtcgccatga gcatcgggtg catcgcaag tgtatcgcc gtggggcaaa tatgactcta 10080
tcgacgcggg cgaccgctac caggtgaaac gcatcaccgt gaaaccgggc gaaggTTTgt 10140
cggtacagat gcattatcat cgCGCggAAC actgggtggt tgtcgccggA acggcaaaag 10200
tcactatcaa cggtgatatac aaactgcttg gtgaaaacga gtccatttat attccgctgg 10260
gggcgatgca ctgcctggaa aaccggggA aaatagattt agaattaatt gaagttcgct 10320
ctggtgcatA tcttgaagaa gatgatgtta ttagatgtta tgatcgctat ggacgaaagt 10380
aatatataat aattatttca gaatttagaaa tgataattat aagtttcgt ctggataaac 10440
aatagatagt atgggttgga aaatatgagt tctttaactt gttttaaagc ttacgacatt 10500
cgcggaaat taggtgaaga actgaatgaa gatatgcct ggcgcattgg tcgcgcctat 10560
ggcgaatttc tcaaaccgaa aaccattgtg ttaggcggtg atgtccgtct caccagcgaa 10620
accttaaaac tggcgctggc aaaaggtta caggatgcgg gcgtcgatgt gctggatatt 10680
ggcatgtccg gcaccgaaga gatttatttc gccacgttcc atctcggtct ggtggcggc 10740
attgaagtta ccGCCAGCCA taatccgatg gattacaacg gcatgaagct ggtgcgcgaa 10800
ggggctcgcc cgatcagcgg tgataccgga ctgcgcgacg tccagcgtct ggcagaagct 10860
aacgactttc ctcccgtcga tgaaaccaaa cgcggtcgct atcagcaaAt caatctgcgt 10920
gacgcTTacg ttgatcacct gttcggttat atcaatgtca aaaaccttac gccgctcaag 10980
ctgggtatca actccggaa tggcgacgcg ggtccggtgg tggacgctat cgaagcccgc 11040
tttaaagccc tcggcgcacc ggtggagttA atcaaagtgc ataacacGCC ggacggcaat 11100
ttcccccaacg gtattcctaa cccgttgctg ccggaatgtc gcgcacgcac ccgcaatgcg 11160

gtcatcaaac acggcgcgga tatggcatt gccttgatg gcgatttga ccgctgttc 11220
ctgtttgacg aaaaagggca gtttatttag ggctactaca ttgtcggcct gctggcagaa 11280
gcgttcctcg aaaaaaatcc cggcgcgaag atcatccacg atccacgtct ctccgtgaac 11340
accattgatg tggtgacggc cgccccggc acgccggtga tgtcgaaaac aggacacgcc 11400
tttattaaag aacgtatgcg caaggaagac gccatctacg gtggcgaaat gagcgctcac 11460
cattacttcc gcgatttcgc ttactgtgac agcggcatga tcccgtggct gctggtcgcc 11520
gaactggtgt gcctgaaagg aaaaacgctg ggcgaactgg tgccgcaccg gatggcggcg 11580
tttccggcaa gcggtgagat caacagaaaa ctggcgacc ctgttgaggc gattaaccgc 11640
gtggaacagc attttagccg tgaggtgctg gcggtgatgc gcaccgatgg catcagcatg 11700
acctttgccg actggcgctt taacctgcgc tcttccaaca cgcgaaccgg ggtgcgcctg 11760
aatgtggaat ctcgcggta tgttcagggt atggtaatcc atactcaaga aatattatca 11820
attttgacgt cataaagaat aagccctgac aagtttagggc ttaattaata tatattttt 11880
ttgaattggg gatttgggt aagattttta atatgttatt taatgtgggt gaattaatgt 11940
tgactggaaa ataataatga gaacgaaaaa agcattacac aactttaaag ttgattttt 12000
aattacttt ttattggtt tgctagggtt ttatattcga actgttttg tttcaaaaat 12060
gggaagtgat attactggag tgatgttact attcacacag ttgacagcat atctcaattt 12120
ggcagaatta ggtattggaa ttgcagctgc cagcgtatta tataaaccgc tcagcgagaa 12180
tgaatacaat aaaataactt acataatatc tttgctctca gtcataataca aatatatatt 12240
tgtgtttgtt ttgattcttgc gcgttggat aggtatctgt atttattact ttattgattc 12300
tgtaaagggtt gtaaatggcg ttttttata ttgggctttg ttcgtttta atacatcggtt 12360
gacatatagt tatgctaaat actccacatt attaactgct aatcagcggt actcagcagt 12420
aagaaaaatt caaggtggcg gaaaagttat aataattgta tttcagatata taattttgt 12480
ctttacgcaa agttcatac tttatttggt agttgagact ttaggtattt tttctcaata 12540
tttgattttt aaaaaataa ttggaaacgg aaatcaataat ctcagtaatg aggtttact 12600
tattgaaagc gataaaacttt tgataaaaaa agaattaaaa ataagaataa aaaatatgtt 12660

cttccataaa ataggtgctg tgcttgcct taatacagac tacctgctt gatcaaaaggtt	12720
tctgacatta agttatgtga caattttgg cagctatatg atggatttc agatagtaac	12780
tgtttgatg tcaagtttg ttaatgctat tactgcagga atggtaatt acttaattaa	12840
taaaagtaat ttagaaattt agggaaattt acgtcaattt tatgtgatat ttatgcctt	12900
tgcaacattc atatcactaa atatgtttt tcttgttaat gatttatcg caaaaatggat	12960
aggtgttaat tatacattaa gtaacaccct agttgcatta atgattgtta acgtattcat	13020
tagtgttgc agggtacccct ctgatataattt aaaaaacgca agtggacatt ttgggtgat	13080
ttattatcca ttattagaag gtgtgctgaa tattacgata tccatcattt tggctatcat	13140
tattggatta cctggcatta ttatagggac aatagtatct aacttaatag taataatgct	13200
tgcgaaacca ttatatctt actctaagtt atttaatctt agaaatccga cgagggttta	13260
ttttgaattt atttctcgcc ctatgtata ttcattatgt gtgattgggg tgagctat	13320
attgcgcgat gaaatataattt cattaaagt aagtacatgg ttggattttt ttaacaagct	13380
actcttagtc tctactccta gcatattggt aatatgtgct attttctcta cgatgtga	13440
ctttagatta ttttcagaa aaattatata tgtgattatg aagaaataaa aatttcgaaa	13500
atgtattaaat cgaaattatg caacgagctt tattttata aatgatatgt gatctttcg	13560
cgaataggag taaggatccg tgtaggctgg agctgcttcg aagttccat actttctaga	13620
gaataggaac ttcggaatag gaactaagga ggatattcat atggataaaag ccgtaagcat	13680
ataagcatgg ataagctatt tatactttaa taagtacttt gtatacttat ttgcgaacat	13740
tccaggccgc gagcattcag cgccgtgatc acacctgaca ggagtatgtatgtccaa	13800
aacagatcgg cgtagtcggt atggcagtga tgggacgcaa cttgcgcctc aacatcgaaa	13860
gccgtggta taccgtctt attttcaacc gttccgtga gaagacggaa gaagtgatttgc	13920
ccgaaaatcc aggcaagaaa ctggttcctt actatacggt gaaagagttt gtcgaatctc	13980
tggaaacgcc tcgtcgcatc ctgttaatgg tgaaaggcagg tgccggcacg gatgctgcta	14040
ttgattccct caaaccatat ctcgataaaag gagacatcat cattgatggt ggtaaacacct	14100
tcttccagga cactattcgt cgtaatcgtg agcttcagc agagggtttt aacttcatcg	14160

gtaccggtgt ttctggcggt gaagaggggg cgctgaaagg tccttctatt atgcctggtg 14220
gccagaaaaga agcctatgaa ttggtagcac cgatcctgac caaatcgcc gccgtagctg 14280
aagacggtga accatgcgtt acctatatgg gtgccgatgg cgcaggtcac tatgtgaaga 14340
tggttcacaa cggtattgaa tacggcgata tgcagctgat tgctgaagcc tattctctgc 14400
ttaaagggtgg cctgaacctc accaacgaag aactggcgca gacctttacc gagtggaaata 14460
acggtgaact gagcagttac ctgatcgaca tcaccaaaga tatcttcacc aaaaaagatg 14520
aagacggtaa ctacctggtt gatgtgatcc tggatgaagc ggctaacaaa ggtaccggta 14580
aatggaccag ccagagcgcg ctggatctcg gcgaaccgct gtcgctgatt accgagtctg 14640
tgtttgcacg ttatatctct tctctgaaag atcagctgt tgccgcatct aaagttctct 14700
ctggtccgca agcacagcca gcagggcgaca aggctgagtt catcgaaaaa gttcgtcgtg 14760
cgctgtatct gggcaaaatc gtttcttacg cccaggcctt ctctcagctg cgtgctgcgt 14820
ctgaagagta caactggat ctgaactacg gcgaaatcgc gaagatttc cgtgctggct 14880
gcatcatccg tgccgcagttc ctgcagaaaa tcaccgatgc ttatgccgaa aatccacaga 14940
tcgctaacct gttgctggct ccgtacttca agcaaattgc cgatgactac cagcaggcgc 15000
tgcgtgatgt cggtgcttat gcagtgacaga acggattcc ggttccgacc ttctccgcag 15060
cggttgccta ttacgacagc taccgtgctg ctgttctgcc tgccgaaacctg atccaggcac 15120
agcgtgacta ttttggtgcg catacttata agcgtattga taaagaaggt gtgttccata 15180
ccgaatggct ggattaa 15197