UCA UNIVERSITÉ Clermont Auvergne

Université Clermont Auvergne
Année 2018

UNIVERSITE CLERMONT AUVERGNE

École doctorale des Sciences de la Vie, Santé, Agronomie et Environnement
Faculté de Medicin UMR454 MEDiS

Nº d'ordre

Thèse

Presenté à l'Université Clermont Auvergne

Pour l'obtention de grade de

Docteur de Biologie Santé
Specialité : Microbiologie Santé

Maricarmen ROJAS LÓPEZ

Identification of cross-protective antigens to develop a vaccine against intestinal pathogenic *E. coli* strains: Special target to enterohemorrhagic *E. coli*

Soutenue le 6 Février 2018

Directeurs de thèse :
M. Mickaël Desvaux (Directeur de Recherche, INRA, France)
M. Fabio Polticelli (Professeur, Universita di Roma Tre, Italie)
Co-encadrants :
M. Roberto Rosini (Chargé de Projet, GSK, Italie)
M. Grégory Jubelin (Chargé de Recherche, INRA, France)

Membres du jury
Président :
Mme. Valérie Livrelli (Professeur, Université Clermont Auvergne, Clermont Ferrand, France)
Rapporteurs:
M. Adam Cunningham (Professeur, University of Birmingham, UK)
M. Stefano Morabito (Professeur, Istituto Superiore di Sanità, Rome, Italy)
Mme. Cristina Ulivieri (Professeur, Università di Siena, Siena, Italy)

GSK Vaccines S.r.l., Siena, Italy.
Institut National de la Recherche Agronomique, Clermont-Ferrand, Site de Theix.

In Memory of my sister Any and my father
(May and September 2016)

Dédicace

Este trabajo es con especial dedicación a mi familia, a mi madre que ha sido mi ejemplo de tenacidad, fortaleza y resiliencia ante la adversidad. A mi hermana Bety, por su gran apoyo y ser el soporte de mi madre en nuestras ausencias. A mi padre y hermana Any que partieron antes de ver realizado este gran proyecto, qué no solo es mío sino de todos nosotros, cómo los echamos de menos.

Dédicace

A ti, dedico también este trabajo. A ti que siempre has estado apoyándome a pesar de la distancia, sin ti esto no hubiera sido ni posible ni lo mismo. A pesar de los años y las distancias siempre sé que estas ahí. Sabes bien que me tuve que ir no solo por mí sino por ti, pero pesar de todo yo te sigo y sí te seguiría dónde sea y cómo sea y hasta al fin del mundo...

Remerciements

A miei amici italiani, che mi hanno mostrato altra definizione di amicizia Ceci, Erika, Elisabetta, per fare il nostro pranzo più piacevole. A Sara, Fede, Marti, Luigi, Trine, Enea, Juliette, Bea, Marco, Carlo, Luigi, Cesira, Dani, Domenico, Monica. Grazie per essere con me, specialmente nelle situazioni più difficile per me e la mia famiglia.
Manuele Martinelli, grazie mille per il tuo aiuto e consigli. Giulia Torricelli, grazie per il tuo aiuto per i genomi sequenced.

Roberto ed Edmondo, grazie mille per tutto, non ci sono parole sufficiente per ringraziare tutto quello che avete fatto per me e per sopportare a questo piccolo uragano che sono io!

Pour le françaises, Cecile et Sonia, vous m'avez fait sentir bienvenue ! You have been a good support for me in Clermont and I know I count you as my friends!

Aux coachs Gui Gou et Quentin pour faire moi plus capable de WOD et lift lourde ! et de votre patience !! Yoan pour être toujours trop souriant et faire moi sentir bienvenue, et à l'entier box ! Vous avez fait ma vie à Clermont plus fácile.

Nat, tú fuiste un gran apoyo cuando llegue a Clermont, pareciera que mis ángeles te habían enviado en el momento oportuno.

Tiago, my dearest Portuguese friend, thank you for being such a nice person and friend, your company was so valuable to me during this time!

Ilham there are not words enough to thank you all the things you have done for me. You have been a big support in the adversity of my life, and a good cellmate. Marc, thank you for bringing that Californian spirit to our lives.

My dear friends around the world I cannot find the proper words to thank you for all your support and your words: Carla Blumenttrit, Lucia Galli, Felipe Del Canto, Dani Olarte, Lidia Fuenzalida, Aneth Zertuche, Roberto Cieza.

Also, to my previous supervisors Maguie Arenas, Ygnacio Martinez, Alfredo Torres and Fidel Zavala who have followed my work but mainly because they have always encouraged me for more.

Roberto Rosini, Mariagrazia Pizza, Marco Soriani, Mickaël Desvaux, Valerie Livrelli, Fabio Polticelli and Gregory Jubelin, you have been such a great guidance in this project. I thank you for trusting me this project but more specially for having the opportunity of working with you.

To my review committee, Cristina Ulivieri, Stefano Morabito and Adam Cunningham, thank you for having the time of reading this work and your valuable comments.

Marco Bargagna, thank you for be an important pilar to coordinate this project.

Thank you to the GSK and INRA team for all the facilities to perform this work.

Abstract

Enterohemorrhagic *E. coli* (EHEC) are a major cause of large outbreaks mainly affecting developed countries. From 1982 to 2002, a total of 350 *E. coli* O157 outbreaks were reported in the United States. EHEC infection causes diarrheal disease often associated with clinical complications like hemorrhagic colitis and hemolytic uremic syndrome (HUS). Although efforts focused on hygiene have been implemented in the food supply chain to reduce the risk of the foodborne *E. coli* O157 infection, outbreaks caused by this pathogen are still common. In addition, antibiotic-based therapy is discouraged for their potential undesirable effect in releasing shiga-toxin from the bacteria. Among non-antibiotic preventing strategies, vaccine development is warranted, still nowadays a licensed vaccine specific for human use against EHEC is not available. In this study, we used the Reverse Vaccinology approach applied on the EHEC O157:H7 genome to select new potential vaccine candidates. We identified a panel of 24 of potential protein antigens and we successfully expressed three of them in Generalized Modules for Membrane Antigens (GMMA) delivery system. GMMA expressing these vaccine candidates resulted to be immunogenic, raising a specific antibody response for two of the selected antigens. In particular, immunization with MC001 candidate was able to reduce intestinal EHEC O157:H7 colonization lowering the bacterial count in feces, colon and ceacum tissues in mice. This candidate was found to be homologue to the *Salmonella* Typhimurium lipid A deacylase enzyme (LpxR) and to our knowledge this study was the first report describing it as vaccine candidate. Also, gene distribution and sequence variability analysis showed that MC001 was mainly present and conserved in EHEC O157:H7 and in some EPEC. Given the high genetic variability among and within these pathotypes, the identification and inclusion of this conserved candidate in a vaccine might cover against major intestinal pathogenic strains. Furthermore, because it has been showed that during the infection process some autotransporters, as MC021 can be reactive, we also analysed molecular determinant with an important role for their proper secretion and folding, namely the autochaperon (AC) domain. It appeared the AC is a common feature of autotransporters but strictly associated with passenger domains exhibiting a β–helix fold. Their exposition at the bacterial cell the surface further position the AC as a potential antigenic target and/or development of new treatments. These findings further provide new research directions for the development of non-antibiotic preventive strategy against InPEC in human but also animal.

Keywords: **Reverse vaccinology; InPEC; Enterohemorrhagic *E. coli*; Vacccine development; GMMA.**

Table of Contents

Abstract ..... - 6 -
Introduction ..... - 8 -
An Overview of *Escherichia coli* ..... - 8 -
Enterohemorrhagic *Escherichia coli* (EHEC) ..... - 9 -
First part ..... - 13 -
Chapter 1 Vaccine Development Against EHEC And Other Examples ..... - 14 -
Introduction ..... - 14 -
1. Development of Stx-based treatments ..... - 16 -
2. Attenuated bacteria and bacterial ghost platforms ..... - 17 -
3. Protein-based vaccines ..... - 19 -
4. Plant-based vaccines ..... - 20 -
5. DNA vaccines ..... - 21 -
6. Polysaccharide-based vaccines ..... - 22 -
7. Improving the adjuvant effect ..... - 23 -
8. Autotransporters ..... - 23 -
9. Membrane vesicles in vaccine development ..... - 24 -
10. Future of vaccine development ..... - 27 -
Chapter 2 Secreted Proteins and Adhesins ..... - 28 -
Introduction ..... - 28 -
Colonization factors of STEC from a secretome perspective ..... - 28 -
1. Type II subtype c secretion systems (T2cSS): Type 4 pili ..... - 29 -
2. Type III secretion systems (T3SS): Injectisome and flagella ..... - 29 -
3. Type IV subtype b secretion system (T4bSS): Conjugative pili ..... - 31 -
4. Type V secretion systems (T5SS): Adhesins by numbers ..... - 31 -
5. Type V subtype a secretion system (T5aSS): SAAT and SPATE adhesins ..... - 32 -
6. Type V subtype c secretion system (T5cSS): Trimeric autotransporter adhesin EhaG .. - 33 -
6. Type V subtype e secretion system (T5eSS): Intimin ..... - 33 -
7. Type VII secretion system (T7SS): Pili by numbers ..... - 34 -
8. Type VIII secretion system (T8SS): Curli ..... - 36 -
Second part ..... - 37 -
Results ..... - 37 -
Chapter 1 ..... - 38 -
Identification of a novel vaccine candidate against Enterohemorrhagic *E. coli* (EHEC) O157:H7 by Reverse Vaccinology approach ..... - 39 -
Chapter 2 ..... - 74 -
Identification of the autochaperone domain in the Type Va secretion system (T5aSS): A prevalent feature of autotransporters with a β–helical passenger ..... - 76 -
Conclusions and Perspectives ..... - 87 -
References ..... - 92 -
Annexes ..... - 117 -
Annexe 1 ..... - 118 -
Intestinal pathogenic *Escherichia coli*: insights for vaccines development ..... - 118 -
Résumé long français ..... - 155 -
Suplementary Material ..... - 175 -

Introduction

An Overview of *Escherichia coli*

Escherichia coli is one of the more common microorganisms in the intestinal microbiota. *E. coli* bacterial colonization begins in the first hours after birth, establishing a symbiotic relationship (Nataro and Kaper, 1998b; Kaper *et al.*, 2004). However, some *E. coli* strains have evolved into pathogenic strains. Pathogenic *E. coli* are classified according to the place or system affected, namely intestinal pathogenic *E. coli* (InPEC) and extra-intestinal pathogenic *E. coli* (ExPEC); the latter comprises uropathogenic *E. coli* (UPEC) and neonatal meningitis *E. coli* (NMEC) (Nataro and Kaper, 1998b; Kaper *et al.*, 2004).

The InPEC strains are subdivided into six pathotypes according to the characteristics of the infection they cause, the extended variety in their genetic and phenotypic features, due to their process of evolution, the mechanism of their pathogenesis, clinical patterns and their interaction with the enteropathotype. These six pathotypes are classified as enteropathogenic *E. coli* (EPEC), enterohemorrhagic *E. coli* (EHEC), enteroinvasive *E. coli* (EIEC), enteroaggregative *E. coli* (EAEC), diffusely adherent *E. coli* (DAEC), and enterotoxigenic *E. coli* (ETEC).

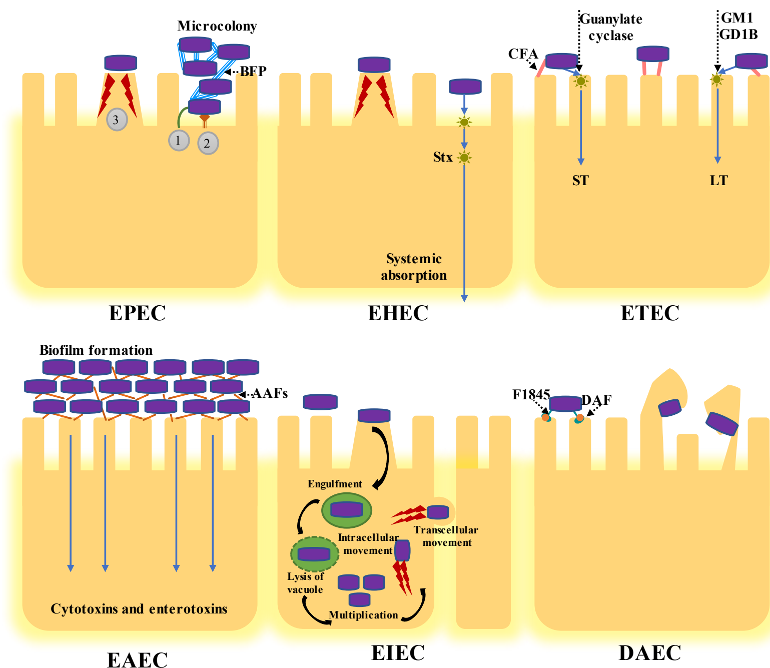


Figure 1. Representation of the six InPEC pathotypes, according to their most studied and representative virulence factors, and their interaction with the enterocyte (Figure based on Kaper *et al.*, 2004). EPEC colonizes small bowel enterocytes, destroys the normal microvillar architecture, and induces the characteristic attaching and effacing (A/E) lesion. Cytoskeletal remodeling occurs during the inflammatory response and diarrhea. 1. Initial adhesion, 2. Protein translocation by type III secretion, 3. Pedestal formation. EHEC colonizes the colon and also induces

the A/E lesion. The hallmark feature of EHEC is the expression of Shiga toxin (Stx). ETEC adheres to small bowel enterocytes provoking watery diarrhea due the secretion of heat-labile (LT) and/or heat-stable (ST) enterotoxins. EAEC adheres to both small and large bowel epithelia through a thick biofilm and elaborates secretory enterotoxins and cytotoxins. EIEC enters colonic epithelial cells, lyses the phagosome and moves through the cell by restructuring actin microfilaments. EIEC might move laterally through the epithelium by direct cell-to-cell spread or might exit and re-enter the basolateral plasma membrane. DAEC causes a characteristic signal transduction effect in small bowel enterocytes displaying the development of long finger-like cellular projections, wrapping around the bacteria. AAF, aggregative adherence fimbriae; BFP, bundle-forming pilus; CFA, colonization factor antigen; DAF, decay-accelerating factor; EAST1, enteroaggregative *E. coli* ST1; LT, heat-labile enterotoxin; ShET1, *Shigella* enterotoxin 1; ST, heat-stable enterotoxin.

Enterohemorrhagic *Escherichia coli* (EHEC)

EHEC is an important pathogen because of its association with several diarrheal outbreaks, mainly following the consumption of water and food contaminated by these bacteria, including lettuce, lucerne, apple juice, and non-pasteurized products, as meat from cattle. The population at risk includes children and the elderly. In the United States in 2014, the incidence per 100,000 population was 690 for EHEC O157 and 445 for EHEC non-O157. The outbreak percentages were: 16% for EHEC O157 and 7% for EHEC non-O157. Regarding EHEC non-O157:H7, their incidence is increasing and becoming a human pathogen as important as EHEC O157:H7 (Stacy M. Crim and Matthew Cartter, 2015). One of the main adverse effects after infection with EHEC is hemolytic uremic syndrome (HUS). This syndrome is mainly driven by expression of Stx and is characterized by a triad of microangiopathic hemolytic anemia, thrombocytopenia, and acute renal failure, with a 10% chance of HUS development in children under 10 years of age. Besides triggering HUS and later renal failure, EHEC is also responsible for strokes (Tarr *et al.*, 2005). Outbreaks have occurred in developed countries, as the United States, Canada, Australia, Japan, France, and others in Europe. Furthermore, Latin American countries have reported EHEC outbreaks and it is a public health problem in Argentina, Brazil and Chile (Nataro and Kaper, 1998b; Kaper *et al.*, 2004) (WHO). (Caprioli *et al.*, 2005; Mejias *et al.*, 2016).

The clinical manifestations of O157 and non-O157 infections are similar: from mild watery diarrhea to bloody diarrhea known as hemorrhagic colitis (HC). The incubation period before the onset of symptoms for O157:H7 is about 3 days (Bell *et al.*, 1994). The main symptoms at the beginning of the infection are diarrhea, and in some cases fever, abdominal cramping, or vomiting (Tarr *et al.*, 2005). HC develops over the following days and it is

recommended that patients be hospitalized, to manage diarrhea and prevent further contagion (Tarr *et al.*, 2005). Usually, patients recover from EHEC infections after a week, but to date we have no way to prevent infection or the development of HUS. Nonetheless, the suggested treatments for EHEC infections are related to the patient's symptoms and signs, as hydration by intravenous fluid administration, renal function and platelet monitoring, but more importantly discouragement of the use of antimotility drugs, pain relief drugs and antibiotics (Holtz *et al.*, 2009).

As for each of the characterized InPEC, EHEC are primarily defined by an array of evidence based on clinical manifestations, *i.e.* clinical symptoms, histological and molecular features (Nataro and Kaper, 1998; Kaper *et al.*, 2004). Based on their incidence, involvement in outbreaks and association with severe disease, EHEC can be classified into 4 classes, namely *(i)* class A corresponds to the O157 serogroup with high virulence, *(ii)* class B to the so-called big six non-O157 serotypes O26:H11, O45:H2, O103:H2, O111:H8, O121:H19 and O145:H28, with moderate virulence, *(iii)* class C to serotypes mainly involved in some sporadic cases with moderate virulence, and *(iv)* class D to serotypes never reported in human infection and thus avirulent (Karmali *et al.*, 2003).

As asymptomatic carriers, cattle are the natural reservoir of EHEC; as an anthropozoonosis, EHEC infection can occur via direct (meat, milk) or indirect (vegetables, water) contamination of food products by animal feces (Rivas *et al.*, 2016). The adhesion and persistence of the bacteria are key features in bacterial pathogenesis. EHEC O157:H7 mainly colonize human colon and carry a pathogenic island (PAI) known as the locus of enterocyte effacement (LEE) (like EPEC), which enables production of the A/E lesion in humans, but this phenomenon is mainly observed in *in vitro* assays. However, the PAI LEE is still a determinant factor for colonization and persistence in other reservoirs (e.g. cattle) and even its evolutionary selection is implied in the selection for pathogenic strains (Nataro and Kaper, 1998a; Coombes *et al.*, 2011; Lewis *et al.*, 2015). Interestingly, the EHEC non-O157:H7 strains do not necessarily carry the LEE PAI. (McWilliams and Torres, 2014; Monteiro *et al.*, 2016).

Virulence determinants in EHEC

Shiga toxins:

Stx is the characteristic virulence factor of EHEC strains, and the most important one in the pathogenesis of HUS. Stx is encoded by a chromosomally integrated prophage. It belongs

to the AB family, and has also been classified into two major types: Stx1with 3 subtypes (a, c, and d), and Stx2 with 7 subtypes (a to g) (Scheutz *et al.*, 2012). The subtypes have been isolated not only from humans but also from other sources. The presence of Stx variants can be combined in EHEC strains; some strains carry one or both Stx, and even combine the subtypes. However, Stx2 is the toxin most associated with severe disease (Boerlin *et al.*, 1999). Overall, Stx causes cellular damage after being internalized into cells by interacting with the Gb3 receptor in Paneth and renal cells, causing the depurination (adenine) of the 28S subunit of the ribosome (rRNA 28S) and stopping protein synthesis following by death of eukaryotic cells. Stx can activate the coagulation cascade, intravascular hemolysis and ischemia (Nataro and Kaper, 1998b; Donnenberg and Whittam, 2001; Thorpe, 2004). The use of antibiotics in the treatment of EHEC infection can first lead to Stx2 production (e.g. fluoroquinolones) (Zhang *et al.*, 2000) and then to cellular damage, either by disrupting the bacteria releasing the toxin or by increasing the production of the toxin by activating the SOS system and stimulation of the lytic cycle of the prophage, e.g. with subinhibitory concentrations of fluoroquinolones and trimethoprim (Kimmitt *et al.*, 2000; McGannon *et al.*, 2010). The resulting secretion of the toxin into the blood stream can lead to worsening of the disease (Pacheco and Sperandio, 2012).

O157 Plasmid (pO157)

The mega plasmid of EHEC, the 92 kb pO157, is very common among the STEC/EHEC O157:H7 strains. It is an F-like plasmid that carries genes encoding important virulence factors, including: *stcE* (a zinc metalloprotease) secreted by the type II secretion system (T2SS); the genes *etp* that encode a T2SS; the operon that includes *hlyA*, for hemolysin; *katP*, which codes for a catalase peroxidase and a serine protease, *espP*, involved in the cleavage of coagulation factor V; the inhibitor of sterase C1 (C1-INH) of the complement viq, it has mucinase activity and is involved in colonization and tissue damage (Burland *et al.*, 1998; Donnenberg and Whittam, 2001; Torres, 2008)

Adherence factors

The mechanism of adherence of *E. coli* O157:H7 and other EHEC serotypes to the epithelial cells in the intestine is one of the crucial steps in pathogenesis. The adherence factor best studied and with major relevance in intestinal colonization *in vivo* and in animal models is the LEE, a 35 Kb pathogenic island organized in 5 operons (*LEE1-5*) and conferring the phenotype of the A/E lesion produced by tight adherence between the Tir and intimin proteins

encoded by the *tir* and *eae* genes, respectively. Furthermore, these adhesins are necessary but not sufficient to produce the A/E lesion *in vitro* and *in vivo*, which is evidence of the need for other factors of adhesion for the initial interaction with the enterocyte (Torres *et al.*, 2002; Torres *et al.*, 2005).

Complete genome sequencing of two EHEC O157:H7 strains indicated the existence of at least twelve non-fimbrial adhesins. The most studied factors for the adherence of this bacterium are: 1) EspA filament, assembled from the expression of the LEE-encoded gene, *espA*; 2) Intimin and Tir, encoded by *eae* and *tir*, respectively; 3) The pO157 product as ToxB, which regulates LEE gene expression; 4) Outer membrane protein A (OmpA); 5) The IrgA homologue adhesin (Iha), homologous to the IrgA in *Vibrio cholera*; and 6) the calcium-binding antigen 43 homologue (Cah), a surface protein (Stevens, 2005; Torres *et al.*, 2005)

Genome sequencing also showed at least sixteen putative fimbrial operons (Perna *et al.*, 2001). The most characterized are 1) type I fimbriae (McWilliams and Torres, 2014); 2) curli (McWilliams and Torres, 2014); 3) F9 fimbriae (McWilliams and Torres, 2014); 4) *E. coli* common pilus (Ecp) (Rendon et al., 2007); 5) type IV pili, hemorrhagic *coli* pilus (HCP) (Xicohtencatl-Cortes et al., 2007 and 2009); 6) ELF and (Samadder et al., 2009); 10) the long polar fimbriae, Lpf1 and Lpf2. (Torres *et al.*, 2002; 2004).

As mentioned above, the molecular pathogenesis of the big six LEE-negative EHEC strains, which are also involved in diarrheal diseases and HUS, is different from that of the LEE-positive strains. The most recent outbreak in Germany was caused by an Stx-producing *E. coli*, but this strain is closer genetically to the EAEC pathotype (Rasko *et al.*, 2011) and to its mechanism of pathogenesis.

The most relevant adhesins found in LEE-negative EHEC are the Lpf2 homolog (LpfO113) (Doughty *et al.*, 2002), STEC autoagglutinating adhesin (encoded by *saa*) (Paton *et al.*, 2001), the STEC autotransporter contributing to biofilm formation (Sab) (Herold *et al.*, 2009) and the *E. coli* immunoglobulin-binding (Eib) protein G (Lu *et al.*, 2006; Merkel *et al.*, 2010).

Due the involvement of Stx in the course of infection and treatment restrictions, different preventive strategies have been implemented and are exemplified in the following sections.

First part

Chapter 1
Vaccine Development Against EHEC And Other Examples

Introduction

The burden of diarrheal diseases caused by several ethological agents is today a worldwide major public health problem that is exacerbated by a lack of preventive strategies such as vaccines to block infections. Besides InPEC, these diarrheagenic bacterial agents include *Cryptosporidium, Campylobacter, Shigella* and *Salmonella*, which are responsible for the vast majority of diarrheal diseases worldwide (MacLennan and Saul, 2014). Mathematical approaches have been developed to standardize the estimation of the burden of diarrheal disease, the etiology and the number of the most affected populations. Specifically, to calculate numbers in the burden of diarrheal disease in children under 5 years old (U5), focused not only on diarrheal diseases but other diseases as pneumonia, the Global Burden of Disease (GBD by Health Metrics and Evaluation, 2010) and Child Health Epidemiology Reference Group (CHERG by World Health Organization) (Kovacs *et al.*, 2015) indexes were implemented. The main goal of these approaches is the estimation of cause-specific disease and morbidity-mortality, but the output yields different numbers due the diverse data included in these methods. In addition, the Global Enteric Multicenter Study (GEMS), a case-control study still ongoing in Africa and Asia, is designed to identify the etiology and population-based burden of pediatric diarrheal disease (Kotloff *et al.*, 2013). Overall, the use of these approaches will allow us to refine the regulations, therapies, and control strategies and to estimate the real numbers of infectious diseases depending on the specific target (Pires *et al.*, 2015).

InPEC contribute significantly to the burden of diarrheal diseases, which are a critical issue in low- and middle-income countries (in Asia, Africa and Latin America). For instance, EPEC and ETEC are mainly endemic in developing countries and in travelers to these countries (Kaper *et al.*, 2004; Torres, 2017), whereas EHEC are responsible for large outbreaks around the world. This pathotype mainly affects developed countries and causes not only diarrheal disease, also clinical complications like hemorrhagic colitis and HUS, which is an increasing problem in Latin American countries like Argentina (Kaper *et al.*, 2004; Pianciola *et al.*, 2016;

Torres, 2017). Similarly, EAEC has also occasionally been involved in diarrheal diseases in developing and industrialized countries (Foster *et al.*, 2015) and in 2011 an EHEC-EAEC hybrid emerged and caused a large outbreak in Europe, with 3816 reported cases, leading to 845 HUS cases and 54 deaths (Brzuszkiewicz *et al.*, 2011; Frank *et al.*, 2011).

Overall, *(i)* the increase in the burden of the disease as traveler diarrhea (ETEC, EAEC), *(ii)* endemic ETEC and EPEC cases in developing countries, *(iii)* infections with EHEC and an increasing number of HUS episodes, *(iv)* the annual cost for the healthcare system, and *(v)* the emergence of antibiotic-resistant strains all point to the need for effective preventive treatment to reduce the burden of diarrheal disease. Although the use of antibiotics is still important in treating such infections, non-antibiotic strategies are either a crucial option to limit the increase in antibiotic-resistant strains, which become a threat to public health (Torres, 2017), or are absolutely necessary for diseases where antibiotic therapies are not recommended, e.g. EHEC infection (Goldwater and Bettelheim, 2012; Rivas *et al.*, 2016). Among non-antibiotic therapies, vaccine development is a strategy of choice but, to date, there is no universal or specific licensed vaccine against InPEC.

For years vaccine development has included different platforms and approaches as (*i)* pathogens attenuated by exposure to different environmental conditions (heat or oxygen) or by multiple passes in culture media (*in vitro*), a method considered as the most ancient and empirical form of vaccination; (*ii)* detoxified toxin forms, like the detoxified version of diphtheria and tetanus toxin; (*iii)* the use of protein-based vaccines, as hemagglutinin from influenza virus or the vaccine for *Bordetella pertussis*; (*iv)* genetically engineered vaccines, which have been the most exploited alternative to vaccine development, in which antigens can be produced in different vectors that reduce toxicity or collateral immunoreactions (allergies to certain components); (*v)* recombinant proteins engineered to be over-expressed, or antigens that are delivered in outer membrane vesicles, as the meningococcal group B vaccine (Mora *et al.*, 2003; Plotkin, 2014).

However, there are several infectious diseases for which these traditional approaches have failed and for which vaccines have not yet been developed. With the advent of whole-genome sequencing and advances in bioinformatics, the field of vaccinology has changed. Approaches like reverse vaccinology, based on the scanning of the annotated complete pathogen genome by bioinformatic prediction of the most likely vaccine candidates, have allowed the identification of promising antigens and the development of a safe broadly protective vaccine against the *Neisseria meningitides* serogroup B pathogen (Pizza *et al.*, 2000;

Giuliani *et al.*, 2006; Feavers and Maiden, 2017). For other pathogens such as *Streptococcus ssp.*, and for extra-intestinal pathogenic *E. coli* (ExPEC), a number of promising antigens have also been identified (De Gregorio and Rappuoli, 2012; Sjoling *et al.*, 2015). In particular, for ExPEC the genome sequence analysis of a neonatal meningitis isolate (NMEC) enabled the identification of 230 potential antigens. The most protective antigens uncovered by that analysis were a broadly conserved adhesin (FdeC) and a conserved secreted zinc metallopeptidase (SslE,), which conferred cross-protection in three different murine models, including intestinal, ascending urinary tract infection and sepsis (Moriel *et al.*, 2010; Nesta *et al.*, 2012; Nesta *et al.*, 2014).

For several years now there has been a marked effort to pursue the development of vaccines against InPEC by various research groups around the world. To date, the main pathotypes considered for vaccine development are EHEC and ETEC, because of their major impact on public health burden. Although EPEC is a major issue in developing countries, no notable advances have been made.

In the case of EHEC vaccines, due the involvement of Stx in the course of infection and treatment restrictions, different preventive strategies, as vaccination, have been implemented. So far, there is no a licensed vaccine for EHEC and several studies are focused on finding suitable vaccine candidates. Some of these studies are exemplified below.

1. Development of Stx-based treatments

Neutralizing the effect of the toxin has been one of the strategies to limit its effect. Antibodies cαStx1 and cαStx2 have been directly engineered against the B subunit of Stx1 and A subunit of Stx2, respectively (Bitzan *et al.*, 2009). Tolerability and the pharmacokinetic profile have been evaluated using chimeric anti-Stx1 and anti-Stx2 antibodies. This combination comprises the variable regions of the murine Stx1-neutralizing or Stx2-neutralizing monoclonal antibodies (mAbs) 13C4 and 11E10, respectively, fused to the human kappa light chain-constant domain sequence and the human immunoglobulin G1 (IgG1) heavy chain constant-domain sequence. These antibodies were well tolerated and safe as antitoxins in healthy human volunteers in a human single-dose study. Although these mAbs neutralize the effect of Stx in mouse experiments, it remains to be shown whether or not they avoid the development of HUS after diarrhea caused by EHEC (Bitzan *et al.*, 2009).

More recently, camelid single antibodies against Stx2 have been tested for the protection they afford against Stx2 (Mejias *et al.*, 2016). Anti-single chain antibodies (VHH)

were obtained with two copies of anti-Stx2B VHH and one anti-serum albumin VHH. This trivalent molecule decreased toxicity in an Stx2 lethal mouse model. Because of its antitoxin effect, VHH has been proposed as an alternative for treatment of the consequences of HUS (Mejias *et al.*, 2016).

2. Attenuated bacteria and bacterial ghost platforms

Gene regulators (global or specific) are important players in the expression, or silencing, of virulence factors in pathogens. In fact, deleting those that promote the expression of specific virulence factors could be useful in attenuating a pathogen. This strategy has been applied to EHEC by deleting the LEE-encoded regulator (Ler). Ler is an important specific regulator that positively regulates the expression of LEE genes involved in the A/E phenotype, as well as genes out of the PAI, including adhesins and genes in the plasmid pO157. An EHEC O157:H7 86-24 strain in which the *ler* and *stx*2 genes (ler/Stx2) have been deleted, but which carries a plasmid that expresses a detoxified version of the Stx1A and Stx2 A subunits, has been used as an attenuated vaccine candidate. These bacteria of reduced toxicity and safe for animal administration were used to immunize mice intraperitoneally. Animals were challenged using the EHEC O157:H7 wild-type strain, which after 6 days was not detected in feces, indicating that the attenuated strain reduces wild-type colonization. Also, in passively immunized suckling offspring born to mothers previously immunized with this vaccine there was a 70% survival rate after challenge with wild-type EHEC bacteria (Liu *et al.*, 2009).

As EPEC is the closest pathotype to EHEC, it has been used as live attenuated vaccine to assess cross-protection among the two pathotypes, but protection against EHEC was only partial. Intragastric immunization with a clinical isolate EPEC O126:H6 and challenge using the EHEC O157:H7 wild-type strain yielded sick mice, but with no dead mice reported. More importantly, it was shown that the EspB and intimin antibodies produced after EPEC vaccination were cross-reactive against EspB and intimin from EHEC (Calderon Toledo *et al.*, 2011).

Using a different delivery system, such as an attenuated *Salmonella* strain, recombinant EspA (300 amino acids of carboxyl-terminal), intimin and the B subunit of Stx2 proteins have been expressed. Mice either orally immunized only or orally immunized and subcutaneously boosted raised similar levels of specific IgG and IgA antibodies against these three antigens. However, intimin-specific antibody levels were higher in mice orally immunized and subcutaneously boosted than in mice orally immunized only. Yet, IgG antibodies specific for

Stx increased a week after the booster vaccination; IgA specific for Stx2B antibodies were only detected in feces and increased even more when boosted. An attenuated *Salmonella* strain harboring the three proteins (intimin, EspA and Stx) showed higher titers of specific antibodies, and specific lymphocyte proliferation, including better protection than in single or double recombinant strain.

Another attenuated bacterial vaccine was constructed to express a γ-intimin variant, encoded by the *eae* gene, in attenuated *Salmonella enterica* serovar Typhimurium χ3987 (Δ*cya*, Δ*crp*, Δ*asd* and H683 Δ*aro* Δ*asd*). Vaccination using these attenuated bacteria expressing intimin was able to increase titers of IgG in serum and IgA in feces, indicating immune responses by the systemic and humoral immune systems. The attenuated bacteria expressing intimin were still detectable in Peyer's patches and spleen, while in feces the colony-forming unit (CFU) count decreased from day 2 to day 10 and later remained constant. In animals immunized with the attenuated *Salmonella*, EHEC O157:H7 shedding post-challenge decreased and IgA and IgG production increased (Oliveira *et al.*, 2012).

A mucosal antigen delivery system used for EHEC vaccine was based on Bacillus Calmette-Guérin (BCG) (a live attenuated strain of *Mycobacterium bovis*), because of its mucosal humoral immune reactivity. A recombinant rBCG-Stx2B expressing the Stx2 subunit B was generated and then used to immunize mice. This recombinant bacillus increased Stx2 IgG in serum to levels that were directly proportional to CFU concentrations. Protection was confirmed when, after challenge with a wild-type EHEC strain, there was a higher survival rate (63%) in immunized mice with higher CFU concentrations than in immunized mice with fewer CFUs of rBCG-Stx2 or null rBCG (Fujii *et al.*, 2012).

Finally, the most recent and studied delivery system for development of a vaccine against EHEC is based on the bacterial ghost (BG) platform. BGs of *E. coli* O157:H7 have been produced by the controlled expression of the X174 lysis gene. This gene produces empty bacterial cell envelopes with the composition of the cell envelope of a living cell, such as lipopolysaccharides (LPS), lipids, peptidoglycans (acting as adjuvants), but which lack the capability to produce infection. These bacteria show an antitoxicity effect on Vero cell cultures, and are safe for administration to mice (Mayr *et al.*, 2005; Cai *et al.*, 2010). Animals orally immunized twice (days 0 and 28) with BG and then challenged at 55 days stopped the shedding of the bacteria after day 3 post-EHEC and showed a rate survival of 93% (Cai *et al.*, 2010). By contrast, in mice rectally immunized against EHEC O157:H7 there was a 100% survival rate after challenge with bacterial shedding until 3 to 5 days after (Mayr *et al.*, 2012).

However, mice that were orally or intragastrically immunized still showed disease symptoms from day 2 to 7 post-challenge, including anorexia, slowing of activity, no stimulus reaction, and convulsions before death (Cai *et al.*, 2010). Data also showed that surviving mice recovered after 7 days and dead mice developed glomerular necrosis and enterocyte effacing. Interestingly, specific IgA and IgG antibody titers were raised in sera and colon from mice immunized twice orally with BG (Cai *et al.*, 2010). In fact, the vaccination of mice with BG is, per se, able to trigger the immune response, observed in Th1/Th2 cell proliferation and higher INF levels in spleen cells, leading to increase titers of IgG in serum and IgA, in both serum and colon samples (Mayr *et al.*, 2005; Cai *et al.*, 2010; Mayr *et al.*, 2012).

More recently, a BG (rSOBG) expressing an exposed Stx chimeric protein comprising the Stx2A and Stx1B subunits (Stx2Am-Stx1B) has been engineered. This rSOBG showed specific IgG and IgA antibody titers to StxA1 and StxB2, and the rate of survival was higher (52%) than with native bacterial ghost-OBGs (12%) when mice were challenged intragastrically with high doses of viable *E. coli* O157:H7. Also, there was no tissue damage in the liver, kidney or intestine of rSOBG-immunized (Cai *et al.*, 2015).

3. Protein-based vaccines

Among several strategies applied to the development of vaccines against, EHEC chimeric protein construction has proven attractive in recent years. A fusion protein comprising Stx2A and the N-terminus of EspA has been tested for its immunoreactivity. Mice subcutaneously immunized with EspA-StxA1 fusion protein showed high titers of IgG antibodies specific to EspA-StxA1 after 7 days. This strong humoral immune response resulted in > 95% survival in mice after a challenge with crude toxin Stx2. Although *in vitro* assays on HeLa cells showed that the anti-EspA-Stx2A1 serum neutralized the action of Stx2 and affected actin rearrangement, it did not prevent the adherence of bacteria to HeLa cells (Cheng *et al.*, 2009).

Another fusion protein constructed with the subunit B from Stx1 and Stx2 and a truncated intimin protein (SSI) increased IgG titers in mice, with specific IgG antibodies peaking at day 14 and IgG specific for Stx2B from the fusion protein decreasing only 54 days after the 3rd immunization. IgG titers were raised by immunization with intimin alone and also with intimin fused with the Stx protein. Immunized mice orally challenged with EHEC O157:H7 88321 showed a 100% survival rate. However, the protection using chimeric vaccines is dependent on the number of immunizations and the bacterial challenge dose (Cai *et al.*, 2011).

Immunization with this SSI chimeric protein avoided pathological damage in the colon and kidney; the antibodies also produced anti-toxin and had an anti-adhesion effect, which was absent in the vaccines using the single proteins (Gansheroff *et al.*, 1999) (Gao *et al.*, 2009). Even though the toxins could have contributed an adjuvant and neutralization effect, the fusion protein did not avoid wild-type adhesion in *in vitro* assays as in other vaccine subunits (Gansheroff *et al.*, 1999). Further, an Stx1B subunit and an enzyme-inactive Stx2A subunit (Stx2Am-Stx1B, SAmB) vaccine induced a Th2-mediated humoral immune response and its typical cytokines, IL4 and IL 10, but a low level of INF-γ. Mice immunized with this chimeric protein were challenged with a lysed EHEC 88321 preparation and showed 93% survival, and even higher rates of survival were obtained when challenging mice with Stx1, Stx2 or Stx1/Stx2. However, disease manifestations were still evident (Gao *et al.*, 2011).

Also, specific peptides have been designed for protection against EHEC. An example is the C terminal region of intimin associated with the A/E lesion that reduced bacterial attachment to Hep-2 cell cultures. Vaccination with this fragment was also associated with protection in mice infected with *E. coli* O157:H7 (Wan *et al.*, 2011). B-cell epitopes of this protein were predicted by structural and antigenicity analysis, and proposed as synthetic vaccine candidates for EHEC (Wan *et al.*, 2011). The best peptide found, KT-12 (KASITEIKADKT) conjugated with KLH, was used to immunize mice by either the subcutaneous (SC) or intranasal (IN) route. Both routes of immunization induced high concentrations of IgG, but SC more so. By contrast, IgA titer was higher with IN immunization. Although this peptide did not fully protect mice infected with the bacteria, it triggered the immune response in them (Wan *et al.*, 2011).

4. Plant-based vaccines

Safety is an important aspect of vaccine development. One approach to reducing the risk of unwanted side effects is the use of plant-based vaccines targeting mucosal immunity. The rationale for using plant cells is based on protecting the antigens from the gastrointestinal tract by plant microencapsulation systems, safe oral delivery and low production cost (Wen *et al.*, 2006; Amani *et al.*, 2009). A toxoid of Stx was engineered (by inactivating subunit A) to be expressed in the *Nicotiana tabacum* (tobacco) NT-1 cell line. Mice were orally immunized by feeding them with these cells expressing the toxoid, or by parenteral immunization and later oral immunization. Mice immunized orally (fed plant extracts once a week for 5 weeks) and mice immunized via the intraperitoneal route and later twice fed NT-1 cells expressing stx2,

showed elevated specific Stx2 anti-IgA in fecal samples, but levels were higher in the orally immunized mice. Furthermore, sera from these mice neutralized Stx2 toxicity in Vero cell cultured, but with greater neutralization using sera from orally immunized mice. The challenge showed that the best protection against the STEC wild-type strain B2F1 (producer of hyper-toxic Stx2d) was by oral immunization. The group immunized intraperitoneally and boosted by feeding showed survival of only 75% (Wen *et al.*, 2006).

Another example of synthetic genes from EHEC was the chimeric gene composed of *espA*, *eae*, and *tir* antigens (EIT). This gene was codon optimized for expression in plant cells and cloned into a plant-expression vector, using CaMV35S (cauliflower mosaic virus 35S) under the control of FAE promoters for tobacco and canola plants. The solubility of the expressed protein was low in this expression system, but it was still possible to produce a soluble product. The EIT protein was used to immunize mice either subcutaneously or orally, and a later challenge with *E. coli* O157:H7 significantly reduced bacterial shedding. These immunized mice showed increased levels of anti-EIT IgG and IgA and reduction of bacterial colonization and histological damage (Amani *et al.*, 2009; Amani *et al.*, 2010; Amani *et al.*, 2011).

5. DNA vaccines

DNA vaccines have been used in order to avoid pathogens or bacterial traces that could lead to the development of disease in vaccine recipients. The construction of a DNA vaccine encoding the entire Stx2 B subunit plus the last 32 amino acid residues of StxA2, including the signal peptides for A and B, was generated by cloning this construct into the pGMS-CSF plasmid encoding the gene for murine granulocyte-macrophage colony-stimulating factor (GMS-CSF). The Stx2AΔAB DNA vaccine expresses a nontoxic Stx2 mutated form (lacking an active site, consisting of a 32-amino-acid A2 and a complete B subunit as two non-fused polypeptides). This vaccine increased titers of IgG antibodies and also conferred protection to immunized mice challenged with native Stx2. In addition, antibodies raised in mice conferred toxin neutralization in Vero cell cultures (Bentancor *et al.*, 2009).

More recently, a selection of prospective DNA vaccine candidates was performed by bioinformatics analysis of EHEC O157:H7: EDL933 and Sakai strain genomes. The vaccines selected included, among others, a putative pilin subunit gene (Z1538), the gene of a T3SS structural protein (*escC*), the C-terminal side of *escC*, and the gene encoding an outer membrane protein of the bacteriophage Bp933W (*iomW*) (Garcia-Angulo *et al.*, 2014; Tapia *et*

al., 2016). Mice were immunized intranasally and then challenged with the wild-type bacteria. In comparison to the entire *escC* gene, only its C-terminal proved to be the best vaccine candidate, by reducing bacterial counts in feces, colon, and cecum, and also by triggering IgG in sera and IgA in feces. The most interesting finding of this study was that the efficacy and immunoprotection of a vaccine candidate depends on its length and how it is presented to the immune system. In fact, the immunoprotection against EHEC was improved when only the C-terminal domain of EscC was used (Garcia-Angulo *et al.*, 2014; Tapia *et al.*, 2016).

Another DNA vaccine candidate used was the lymphocyte inhibitory factor A-/EHEC factor for adherence-1 gene (*lifA*/*efaA*). This gene encodes a 360 kDa toxin mainly found in non-O157 EHEC strains and associated with LEE strains. It is exposed on the surface of EPEC bacteria and may play a role in colonization and adhesion by mucosal immunity regulation. This gene was originally found as a truncated form in EHEC O157:H7 EDL933 and annotated as *efa-1'*. EHEC carrying the truncated form showed reduced adhesion to human colon cells, showing that Efa-1' protein still has a role in adhesion. Furthermore, mice vaccinated with this *efa-1'* showed IgM, IgG and IgA. Intranasal immunization using pVAX-efa1 showed higher levels of antigen-specific mucosal IgA in nasal and bronchoalveolar lavages. In animal challenge, pVAXefa-1 reduced EHEC colonization in mice (Riquelme-Neira *et al.*, 2015).

6. Polysaccharide-based vaccines

Polysaccharides in conjugate vaccines against *Haemophilus influenzae* type b, pneumococcal and meningococcal bacteremia, and meningitis have been successfully used for vaccine development. *E. coli* isolates produce two serotype-specific surface polysaccharides, namely the LPS O antigen and the capsular polysaccharide K antigen. Variations in structures of these polysaccharides give rise to ~170 different O antigens and ~80 K antigens (Whitfield, 2006). Immunization with an O-specific polysaccharide of *E. coli* O157:H7 led to a significant increase of IgG against LPS. *E. coli* O157:H7 O-specific polysaccharide conjugated to recombinant exotoxin A of *P. aeruginosa* (O157-rEPA) and administered to 2- to 5-year-old children led to a >4-fold increase of serum IgG after the first week of immunization. Anti-LPS IgG increased >8 fold in serum at week 6 and 20-fold at week 26, when there was no difference among the groups receiving one dose or two doses, but was >4-fold higher than in the pre-immune sera. The serum samples had an antibacterial activity correlated with the anti-LPS IgG antibody titers. More importantly, vaccinated children had mild to non-collateral reactions to the vaccine. This prospective vaccine proved to be a good candidate because of its safety and

immune reactivity (Konadu *et al.*, 1994; Konadu *et al.*, 1998; Konadu *et al.*, 1999; Ahmed *et al.*, 2006).

7. Improving the adjuvant effect

In vaccine development, the adjuvant is important as it enhances the immunogenicity and reactivity of the antigens in the host immune system. Vaccination with EspB or the C-terminal of γ-intimin (280 amino acids) co-administered with the MALP-2 adjuvant (TLR6 agonist) enhanced IgG specific antibodies after a first intranasal immunization, in contrast to vaccination without MALP-2, which increased IgG titers after a second boost. Co-administration of intimin and EspB result in higher titers of sIgA in bronchoalveolar lavage, as did intranasal immunization, while IgG levels in the intestine were always lower with MALP-2 as an adjuvant (Cataldi *et al.*, 2008).

The adjuvant effect has also been implemented by the combined expression of toxins and antigens at the same time, as a specific antigen combination. An example of this method, a chimeric protein of Tir, Stx1B, Stx2B and zonula occludens toxin (Zot), has been used in mice to test its immunoreactivity. This Zot protein was used as an adjuvant element in the chimeric construct. This toxin is encoded onto a filamentous bacteriophage CTXϕ of *Vibrio cholera*, which increases the permeability of mucosa by binding to enterocytes and affecting the tight junctions. It does not damage the tissue and its effect is reversible. Intranasally immunized mice showed better immune responses than subcutaneously immunized mice: IgA and IgG response against the chimeric protein and reduced the bacterial shedding in feces post-EHEC wild-type challenge. Comparison of the adjuvant effect of Zot from the chimeric protein Stx2B-Tit-Stx1B-Zot against a chimeric protein lacking Zot protein has shown a better protection level in the presence of Zot, even though there was no specific IgG titer against it (Zhang *et al.*, 2011).

8. Autotransporters

Immunoproteomic analysis has also been employed to identify antigens that generate an immune response in the host during the process of infection. For this type of approach, reacting human or mice sera obtained after ETEC infections enable the identification of immune reactive molecules in culture supernatant, outer membrane, and outer membrane vesicle preparations by matrix-assisted laser desorption ionization–time of flight mass

spectrometry (MALDI-TOF MS). In this analysis, hypothetical proteins homologous to other pathovars (UPEC, APEC, Crohn's disease–associated isolates) and pathogens (*Vibrio cholera* and group A *streptococci*) were identified, suggesting their role in pathogenesis. In addition, autotransporters (AT), such as EtpA, Antigen 43 and TibA, were also found to be reactive during the infection process (Roy *et al.*, 2010b).

In silico tools have been used to identify other remarkable virulence factors as immunogenic molecules. Autotransporters, which are important in biofilm formation, such as autotransporter (pAT), antigen 43 and EatA, have been identified exclusively in ETEC or other pathotypes, but are absent in commensal strains. Recombinant Ag43 and pAT autotransporters increased fecal IgA and provided relative protection against ETEC intestinal colonization in immunized mice. Interestingly, these autotransporters are also recognized by sera from patients with ETEC diarrhea, confirming their expression during ETEC infections (Harris *et al.*, 2011).

9. Membrane vesicles in vaccine development

Outer membrane vesicles have been largely employed as an antigen delivery system. Native OMV (NOMV) are naturally released from Gram-positive and Gram-negative bacteria. Genetically engineered bacteria can increase the amount of membrane vesicles in the media. More specifically, mutations in proteins as Braun's (murein) lipoprotein (Lpp) and proteins from the Tol-Pal system (TolA, B, Q and R) can lead to over-production of membrane vesicles, which are called generalized modules for membrane antigens (GMMA) (Berlanda Scorza *et al.*, 2012). The clear distinction concerning NOMV and GMMA is the way the membrane vesicles are extracted/obtained. Overall, the membrane vesicles approach to vaccination has proven attractive, because of the benefits provided as protection attributed to the presence of pathogen-associated molecular patterns (PAMPs), which are essential in the innate immune response and in promoting adaptive immune responses, and for the benefits of adjuvancy promoted. Hence, GMMA and NOMV represent the most attractive approach in vaccine development, for the expression of heterologous surface antigens, not only to amplify protection against different pathogens, but also to increase the antigenicity of such antigens. Some examples of the use of NOMV and GMMA in vaccinology are considered below.

Native OMV from *V. cholera* used to immunize female mice by the intranasal, intragastric, or intraperitoneal route induced specific, high-titer antibodies of similar levels against diverse antigens present in the NOMV. These antibody titers were stable over a long period (3 months). The induction of Ig isotypes depended on the route of immunization:

immunoglobulin A was induced to a significant level by mucosal immunization, like the intranasal route, which generated the highest titers. In subsequent experiments, offspring born to immunized female mice that were orally challenged with *V. cholerae* 30 and 95 days after the last immunization displayed protection against colonization with *V. cholerae* in both challenge periods, thus validating the development of NOMV-based vaccines (Schild *et al.*, 2008).

In other models, NOMV has been tested for protection against *Burkholderia pseudomallei*. The observations made by Morici's lab were that NOMV conferred protection against lethal aerosol, systemic and sepsis infection in BALB/c mice, as well as a humoral immune response to protective protein and polysaccharide antigens, without any associated toxicity or reactogenicity in non-human primates, both cases caused by *B. pseudomallei* (Nieves *et al.*, 2014; Petersen *et al.*, 2014). More specifically, the immune response in a mouse model was apparent from the induced specific IgG and IgM antibody response. Also, functional serum activity in *in vitro* assays of bacterial killing and passive transfer of *B. pseudomallei* NOMV immune sera protected naive mice against a subsequent challenge. Proposing NOMVs as an efficacious multivalent vaccine strategy against melioidosis (Nieves *et al.*, 2014).

The NOMV employed as a delivery system showed that serine proteases as heterologous proteins expressed on the surface of NOMV from *E. coli* can be used as prospective vaccines. The HtrA serine protease from *Chlamydia muridarum* DO (antigen that confers strong humoral and cellular responses in mice and humans) was fused to the OmpA leader sequence to lead it on *E. coli* NOMV. The sera recovered from mice immunized with NOMV having the HtrA (CM rHtrA-OMV) neutralized chlamydial invasion of LLC-MK2 cells *in vitro*. During this study, it was demonstrated that HtrA delivery through NOMV, where it appears to assume a conformation similar to that of the native one, enhanced immune response and functional antibodies more than the recombinant protein version of HtrA. This difference was associated with a diverse pattern of protein epitope recognition. According to this study, the epitope recognition profile of anti-CM HtrA-OMV antibodies was similar to that induced in mice during chlamydial infection (Bartolini *et al.*, 2013).

As noted above, the OMV are very attractive for the production of multivalent vaccines because of their built-in adjuvanticity and for their potential use as carriers of heterologous antigens. The antibody response induced by OMV from *E. coli* expressing antigens as Group A *Streptococcus* (GAS) Slo, SpyCEP, Spy0269 and Group B *Streptococcus* (GBS) SAM_1372 fused to a OmpA signal sequence, was assessed. The immunogenicity and protective activity

showed not only high functional antibody titers but also protection against GAS lethal challenge in mice by using Slo-OMV and SpyCEP-OMV as vaccines (Fantappie *et al.*, 2014)

Another example of the use of the NOMV system was reported more recently by Leitner *et al.* in the development of a vaccine against ETEC. NOMV from ETEC and *Vibrio cholerae* were used in two different ways. The first one used an ETEC mutated in lipid A acyltransferase (msbB) and labile toxin subunit A (eltA) (EΔmsbBΔeltA) to decrease the toxicity of OMV. The OMV generated from this mutant were used to immunize adult mice and test immunoreactivity. The detoxified OMV yielded higher titers of IgG1, IgM, and IgA, in comparison to the OMV from a wild-type strain. Offspring born to mothers immunized with the OMV-EδmsbBΔeltA and passively immunized by suckling were challenged with wild-type ETEC, and displayed low colonization levels. In the second part of this approach, antigens from ETEC, adhesins FliC and CFA/I were expressed as heterologous proteins in OMV from *Vibrio cholerae* with mutations in msbB and cholera toxin subunits A and B (ctxAB) and the major flagellin FlaA (flaA). As in the previous test with the OMV from EΔmsbBΔeltA, the immune response displayed the same pattern in adult mice and the same pattern of colonization in neonatal mice. Even though they showed no differences in immunoreactivity pattern or protection, this approach showed the benefit of using OMV in vaccine development, not only for ETEC, but also for other pathogens. (Leitner *et al.*, 2015).

Also, membrane vesicles (NOMV or GMM) have been used to express and deliver heterologous proteins. NOMV from *E. coli* were used to express and deliver one, two, or three *Mycobacterium tuberculosis* antigens: ESAT6, Ag85B, and Rv2660c, fused to hemoglobin protease (Hbp) autotransporter from *E. coli*. The aim of fusing these *M. tuberculosis* antigens to the Hbp was to target to the surface of *Escherichia coli* NOMV, which was later expressed in a hypervesiculating attenuated *Salmonella enterica serovar* Typhimurium and mutant in *tolR* and *tolA* (GMMA) upon fusion to Hbp. More importantly, the GMMA from Salmonella decorated with Ag85B, a potent T cell antigen presented in cells *in vitro*, processed and presented an epitope functionally recognized from this protein by Ag85B-specific T cell hybridomas. This platform was shown to be efficient in exhibiting heterologous antigens individually or combined by detection of antigen-specific immune responses upon vesicle-mediated delivery (Daleke-Schermerhorn *et al.*, 2014).

Shigella sonnei strains genetically modified (*Δ tolR:: kan Δ virG:: nadAB Δ htrB:: cat*) to produce penta-acylated LPS with reduced endotoxicity and to maintain the virulence plasmid encoding the immunodominant O antigen component of the LPS, were used to immunize

rabbits and mice. The data from mice demonstrated substantial anti-LPS antibody levels, but most importantly lower reactogenicity. The modified *S. sonnei* GMMA vaccine has been employed in Phase 1 dose-escalation clinical trials (Gerke *et al.*, 2015) To prevent invasive nontyphoidal *Salmonella* (iNTS, *Salmonella enterica* serovars Typhimurium and Enteritidis), the GSK Vaccines Institute for Global Health (GVGH) has extensively employed GMMA for vaccine development. GMMA with modifications to their LPS, from *Salmonella*, as lipid A modification reduces potential reactogenicity *in vivo,* are under study in vaccine development. GMMA with penta-acylated lipid A (by mutating *msbB* and *pagP*), reduced cytokine release from human peripheral blood monocytes in comparison to GMMA with wild-type lipid A, given that GMMA from *S. enteritidis* Δ*msbB* and Δ*pagP* had a slightly higher stimulatory potential than those from *S. typhimurium* (Rossi *et al.*, 2016). GMMA from *S. typhimurium* and *S. enteritidis*, both Δ*msbB* and Δ*pagP*, showed similar immune stimulation to that of *Shigella sonnei* GMMA, which is currently in phase I clinical trials as mentioned above (Gerke *et al.*, 2015).

10. Future of vaccine development

EHEC infections and associated diseases are also related to other 6 STEC serotypes, O26:H11, O45:H2, O103:H11, O11, O121:H19, O145, than the canonical serotypes O113:H21 and O157:H7. Treatments and prevention, like vaccine development, should also target these strains. More recently, new approaches have been implemented to find new antigens. For example, an immunoproteomics analysis developed by a Chilean group detected antigens with some immune reactive effect in patients infected with any of these STEC serotypes. It was mainly outer membrane proteins, like OmpT and Cah, that were reactive in HUS patients. The genes encoding these proteins are widely represented among *E. coli* pathotypes and commensal strains. Other detected proteins included FliC, Ag43 (ETEC), NmpC, OmpF, OmpC, OmpA, Hek, EF-Tu, and L-asparaginase II. (Montero *et al.*, 2014)

Chapter 2
Secreted Proteins and Adhesins

Introduction

As part of vaccine development, and as exemplified above, the main targets are all those exposed structures suitable for recognition as secreted proteins, fimbrial and afimbrial adhesins, colonization factors or outer membrane proteins. For these reasons, the study and description of them is important.

Basically, pathogenic organisms differ from their nonpathogenic counterparts by the encoding of specific virulence factors that are secreted at the bacterial cell surface or into the immediate environment (Finlay and Falkow, 1997). To date, nine protein secretion systems, the Type I to Type IX secretion systems (T1SS to T9SS), have been uncovered in Gram-negative bacteria (Chagnot *et al.*, 2013). Considering the wealth of systems and effectors, the secretome is a powerful concept to apprehend protein trafficking as a whole in living cells by considering both the secretion routes and their cognate secreted proteins (Tjalsma *et al.*, 2000; Antelmann *et al.*, 2001). While the LEE is a key and prominent molecular determinant in pathogenicity, neither EHEC nor STEC all possess the LEE, suggesting that they possess additional virulence and colonization factors (Paton *et al.*, 2001). In terms of the framework for STEC, including related enteropathotypes such as EHEC, the complement of the secretome associated with the expression of exposed structures is described here, especially those that are involved in colonization. Colonization processes are essential in human infection, but also for bacterial multiplication in animals and contamination of foodstuffs. In order to highlight the molecular mechanisms responsible for their correct expression and subcellular localization, the effectors are considered here from the perspective of a protein secretion system.

Colonization factors of STEC from a secretome perspective

Protein secretion is an essential cellular function present in all living cells. In bacteria, it is involved in a huge diversity of important functions in bacterial physiology, from adaptation and response to environmental cues, adhesion and biofilm formation, to pathogenicity and virulence (Chagnot *et al.*, 2013). While secretion refers to the active transport from the interior to the exterior of the cell compartment, secreted proteins have three main possible fates in diderm-LPS bacteria (archetypal Gram-negative bacteria): (i) remain associated with the

bacterial outer membrane (OM), (ii) be released into the extracellular milieu, or (iii) be injected into a target cell (either a eukaryotic or bacterial cell) (Gerlach and Hensel, 2007). Among the nine protein secretion systems in diderm-LPS bacteria (Chagnot *et al.*, 2013), T1SS to T8SS can be involved in surface colonization of STEC (Figure 1). Respective to those systems, the secreted proteins involved in the surface colonization process, namely adhesion and biofilm formation, are reviewed in STEC (Table 1). In addition, some outer membrane proteins (Omp) assembled by the β-barrel assembly machinery (Bam) complex (Knowles *et al.*, 2009), namely IrgA homologue adhesion (IhA) (Tarr *et al.*, 2000; Herold *et al.*, 2009b) or OmpA (Torres and Kaper, 2003), can also participate in the colonization process in STEC.

1. Type II subtype c secretion systems (T2cSS): Type 4 pili

T2SS, or the secretion-dependent pathway (SDP), is responsible for the secretion of proteins first exported via the Sec or Tat systems (Voulhoux *et al.*, 2001). While T2SS is primarily involved in the secretion of free soluble extracellular proteins, namely through the subfamilies Xcp, *i.e.* T2SS of subtype a (T2aSS), and Hxc, *i.e.* T2SS of subtype b (T2bSS), it can also be involved in pili assembly (Durand *et al.*, 2011). Indeed, the system responsible for subunit secretion, assembly and biogenesis of the type 4 pili (T4P) is a distant homologue to the paradigm Xcp T2aSS, and actually corresponds to the Type 2 subtype c secretion system (T2cSS) (Ayers *et al.*, 2010; Chagnot *et al.*, 2013). In pathogenic *E. coli*, T4P are involved in bacterial virulence, including the colonization process, especially host cell adherence, biofilm formation, bacterial aggregation and twitching motility (Craig *et al.*, 2004). In STEC, the T4P named hemorrhagic coli pili (HCP) is composed of a 19-kDa pilin subunit encoded by the *hcpA* chromosomal gene, also called prepilin peptidase-dependent (*ppdD*) in *E. coli* K-12. In EHEC O157:H7 EDL933, HcpA is directly involved in cell adherence and in the ability to invade human and bovine epithelial cells. HCP also contributes to biofilm formation due to its ability to bind some extracellular matrix (ECM) proteins, especially fibronectin and laminin (Xicohtencatl-Cortes *et al.*, 2009). In some EAEC, T4P are encoded on conjugative plasmids (thus encoding a T4bSS in addition) and contribute to bacterial adherence to epithelial cells and abiotic surfaces, as well as biofilm formation (Dudley *et al.*, 2006).

2. Type III secretion systems (T3SS): Injectisome and flagella

The injectisome and the flagellum are assembled by two phylogenetically distinct but homologous Type 3 secretion systems (T3SS) of subtypes a and b, *i.e.* T3aSS (injectisome system) and T3bSS (flagellar system); respectively (Desvaux *et al.*, 2006b; Diepold and

Armitage, 2015). Among InPEC, T3aSS is the hallmark of EPEC, but is also present in some but not all EHEC (Galan and Wolf-Watz, 2006). T3aSS consists of a double-membrane embedded nanomachine and promotes the delivery of bacterial effectors to the cytoplasm or plasma membrane of target eukaryotic cells, where they can modulate or subvert a large variety of host cell mechanisms, but also promote bacterial invasion and colonization (Cornelis, 2006). The injectisome formed by T3aSS is a needle structure, a multiring complex that spans the bacterial envelope and plasma membrane of the host cell providing a continuous path to the cytosol of the infected host cell. In EPEC, T3aSS is directly involved in cell adhesion and pedestal formation, resulting in characteristic histopathological A/E lesions (Wong *et al.*, 2011). Contrary to what is sometimes wrongly assumed in some of the scientific literature, such lesions are never observed from clinical samples after EHEC infection (Kelly *et al.*, 1987; Nataro and Kaper, 1998a; Lewis *et al.*, 2015). Nonetheless, A/E lesions are frequently observed *in vitro* from EHEC-infected epithelial cell cultures. The T3aSS structure is encoded by genes located in the LEE pathogenicity island (McDaniel *et al.*, 1995). Among all the proteins secreted by the injectisome T3aSS, Tir (translocated intimin receptor) is the primary molecule associated with intimate bacterial interaction with the epithelia and A/E lesions (Donnenberg *et al.*, 1993). Tir, encoded by the *espE* gene located in the LEE, is secreted and injected into the host cell by the injectisome T3aSS, and later localized at the plasma membrane of the infected host epithelial cell. This bacterial protein exposed at the host cell surface then acts as a receptor for direct and specific binding of the bacterial cell surface-exposed intimin. By protein-protein interaction, Tir activates the recruitment and rearrangement of the host cell cytoskeletal actin, which rearranges and results in the formation of a characteristic pedestal structure (Liu *et al.*, 1999b). Furthermore, the injectisome plays a role in adhesion to plants, with a marked tropism for the stomata (Shaw *et al.*, 2008; Berger *et al.*, 2010). Besides adhesion to the host cell, T3aSS also participates in invasion capability in EIEC (Schroeder and Hilbi, 2008; Croxen *et al.*, 2013).

While T3aSS is only present in a subset of InPEC, namely EPEC, EIEC and some EHEC, the flagellum is quite ubiquitous across *E. coli*. In this species, the flagella are peritricheous and form a ponytail when in motion (DiLuzio *et al.*, 2005). The different components of the flagellum are secreted by T3bSS, namely the hook-filament junction protein, the filament-capping protein, and flagellin, the major subunit of the filament (Macnab, 2004; Altegoer and Bange, 2015). Of course, flagella are primarily involved in cell motility, especially swimming and/or swarming in *E. coli* depending on environmental conditions (Turner *et al.*, 2010). Besides coordinated movement at surfaces contributing to colonization, that is swarming,

flagella participate in adhesion and invasion by providing motility towards surfaces or target cells (Haiko and Westerlund-Wikstrom, 2013). In EHEC, flagella play a role in adhesion to epithelial cells, but are down-regulated after contact with the epithelium and so just initiate the early stages of the adhesion process (Mahajan *et al.*, 2009). Following vaccination targeting the flagellin FliC, the colonization of EHEC in cattle was significantly decreased, providing evidence of the importance of flagella in host gut colonization (McNeilly *et al.*, 2008).

3. Type IV subtype b secretion system (T4bSS): Conjugative pili

In terms of protein secretion, T4SS can be divided into two subtypes. T4aSS includes effector translocator systems homologous to the prototypical VirB/D4 complex extensively investigated in *Agrobacterium tumefaciens* which is notably involved in the injection of effectors into infected host cells. T4bSS comprises conjugation machines homologous to the prototypical F-episome conjugal transfer (Tra) system, notably involved the transport of nucleoprotein complexes (Christie and Vogel, 2000; Alvarez-Martinez and Christie, 2009; Christie *et al.*, 2014). While no T4aSS has been described so far in *E. coli*, conjugative plasmids can be present (Lawley *et al.*, 2003). Besides pO157 (Lim *et al.*, 2010), plasmid profiling in different *E. coli* O157:H7 strains revealed the presence of numerous additional plasmids, which greatly vary in size and number (Ostroff *et al.*, 1989; Paros *et al.*, 1993). Among them, pO157-Sal was identified and demonstrated to be a novel conjugative plasmid of the Incl family in STEC (Fratamico *et al.*, 2011; Wang *et al.*, 2011). This plasmid contains the full set of *tra* genes and thus encodes a T4bSS. From what is known of the F episome in *E. coli* K-12, such a plasmid could further contribute to bacterial colonization (Ghigo, 2001). Conjugative F pili assembled by T4bSS induce biofilm formation by improving the adhesion capability of bacterial cells. Horizontal transfer of the F episome to siblings within the biofilm further increases the proportion of transconjugant cells and further expands the colonization propensity of the bacterial population (Ghigo, 2001; May and Okabe, 2008; May *et al.*, 2010).

4. Type V secretion systems (T5SS): Adhesins by numbers

T5SS can be categorized into 5 subtypes, (i) the autotransporter system (subtype a), (ii) the two-partner secretion pathway (subtype b), (iii) the trimeric autotransporter system (subtype c), (iv) the hybrid autotransporter system (subtype d), and (v) the inverted autotransporter system (subtype e). Broadly speaking, T5SS refers to protein secretion via an OM pore formed by a β-barrel, the secreted proteins being first exported via the Sec translocon. Except for T5dSS, all four other subtypes can secrete proteins involved in surface colonization. Although T5bSS is

present in *E. coli* 0157:H7 (Choi *et al.*, 2007), the exoprotein O157:H7 two-partner protein A (OtpA) does not display any adhesive properties towards intestinal epithelial cells and the possible contribution of this secretion system and cognate secreted proteins to surface colonization remains to be established.

5. Type V subtype a secretion system (T5aSS): SAAT and SPATE adhesins

Autotransporters are single polypeptides that can drive their own secretion through the OM via a C-terminal translocator forming a β-barrel in the OM allowing translocation of the N-terminal passenger domain. The main function of T5SS is to secrete virulence factors, but it also participates in cell-to-cell adhesion (Desvaux *et al.*, 2004b). The self-associating autotransporters (SAATs) are cell surface-exposed outer membrane proteins (Omps) encompassing adhesin diffuse adherence (AIDA), enterotoxigenic invasion locus b protein A (TibA) and antigen 43 (Ag43) (Benz and Schmidt, 1989; Klemm *et al.*, 2006) autotransporter family. Besides autoaggregation, SAATs are actively involved in bacteria-host interaction and biofilm formation (Sherlock *et al.*, 2004). Five autotransporters of the AIDA family have been identified in *E. coli* O157:H7 and are called enterohemorrhagic *E. coli* autotransporters (Eha) (Wells *et al.*, 2008). EhaA overexpression in *E. coli* K-12 appears to promote autoaggregation, biofilm formation and adhesion to bovine primary epithelial cells from the terminal rectum (Wells *et al.*, 2008). EhaB specifically binds to some ECM proteins, especially laminin and collagen I (Wells *et al.*, 2009). Similarly, EhaJ mediates specific adhesion to collagens I, II, III and V as well as fibronectin, fibrinogen and laminin (Easton *et al.*, 2011). While EhaD and EhaJ also promote biofilm formation, the function of EhaC in colonization remains to be established (Wells *et al.*, 2008). Interestingly, glycosylation of EhaJ is of importance for biofilm formation, but not for binding to the ECM (Easton *et al.*, 2011). While no TibA homolog has been reported in STEC, *E. coli* O157:H7 encodes an Ag43 homolog called calcium-binding antigen 43 homologous (Cah) that promotes cell autoaggregation, but is apparently not involved in adhesion to host cells (Torres *et al.*, 2005a). The serine protease autotransporters of *Enterobacteriaceae* (SPATEs) constitute another autotransporter family generally released into the extracellular milieu and primarily show proteolytic activity against various substrates, but also adhere to some of them (Henderson *et al.*, 2004). In *E. coli* O157:H7, extracellular serine protease plasmid-encoded (EspP) is directly involved in biofilm formation, but also adherence to epithelial cells (Dziva *et al.*, 2007; Puttamreddy *et al.*, 2010). Recently, EspP was shown to oligomerize to form megastructural ropes, which possess adhesive and cytopathic activities in host epithelial cells (Puttamreddy *et al.*, 2010; Xicohtencatl-Cortes *et al.*, 2010).

In STEC O26, protease secreted by STEC (PssA) was shown to participate in intestinal colonization of calves (van Diemen *et al.*, 2005). In STEC O113, STEC autotransporter contributing to biofilm formation (Sab) is involved in adherence to abiotic surfaces and epithelial cells (Herold *et al.*, 2009a). The high variability in the presence of those different autotransporters in STEC may significantly contribute to differences in colonization abilities and even modulate virulence. For example, the identification of Sab in LEE-negative STEC O113:H2 contrasts with its absence from a LEE-positive strain collection and suggests an alternative way to adhere to the host cells for strains defective in their ability to induce A/E lesions (Herold *et al.*, 2009a).

6. Type V subtype c secretion system (T5cSS): Trimeric autotransporter adhesin EhaG

T5cSS corresponds to autotransporters formed upon homotrimerization. All trimeric autotransporters characterized to date are exposed to the bacterial cell surface and play a role in adhesion (El Tahir and Skurnik, 2001; Linke *et al.*, 2006; Szczesny and Lupas, 2008). The EHEC adhesin G (EhaG) is present in diarrheagenic *E. coli*, from EHEC, ETEC, EPEC, EAEC to EIEC, and is highly prevalent in STEC (Totsika *et al.*, 2012). Upon overexpression in *E. coli* K-12, EhaG from *E. coli* O157:H7 was shown to mediate autoaggregation resulting in the formation of strong biofilm. In addition, adhesion to intestinal epithelial cells and specific binding to collagens I, II, III, and V as well as to laminin, fibronectin and fibrinogen were reported (Totsika *et al.*, 2012). STEC autoagglutinating adhesion (Saa) is involved in adhesion to intestinal epithelial cells and has been identified in various LEE-negative STEC (Paton *et al.*, 2001). On the other hand, *E. coli* immunoglobulin-binding protein G (EibG) contributes to a chain-like pattern of adhesion to human epithelial cells, also in LEE-negative STEC (Lu *et al.*, 2006; Merkel *et al.*, 2010).

6. Type V subtype e secretion system (T5eSS): Intimin

T5eSS refers to inverted autotransporters in the sense that the translocator is located at the N-terminal instead of the C-terminal region of the monomeric autotransporter (Oberhettinger *et al.*, 2012; Leo *et al.*, 2015; Oberhettinger *et al.*, 2015). Intimin from EPEC and EHEC is a prototypical member of this Omp family. Along with T3aSS, intimin is encoded by the *eae* gene in the LEE and, as already mentioned above, interacts specifically with Tir, resulting in intimate attachment of the bacteria to the host-cell surface, pedestal formation and A/E lesions (Schmidt, 2010). While five alleles (α, β, γ, δ and ε) have been reported for *eae* with a total of

27 variants (Adu-Bobie *et al.*, 1998; Oswald *et al.*, 2000), these intimins appear to be functionally interchangeable (Liu *et al.*, 1999b; Mallick *et al.*, 2012). Besides Tir, the binding of intimin to alternative receptors such as β_1 integrins or nucleolin remains unclear (Liu *et al.*, 1999a; Leo *et al.*, 2015). Nonetheless, intimin can also contribute to intestinal colonization independently of its ability to bind to Tir (Mallick *et al.*, 2012).

While no invasin has been reported to date in STEC (Leo *et al.*, 2015), the *E. coli* adherence factor FdeC was recently uncovered as another representative of T5eSS and is widely distributed in *E. coli*, including InPEC, *e.g.* STEC and EHEC (Nesta *et al.*, 2012; Easton *et al.*, 2014). FdeC binds to human epithelial cell and exhibits affinity for some ECM components, notably collagens I, III, V and VI (Nesta *et al.*, 2012). Its contribution to kidney and bladder colonization is associated with the propensity of UPEC bacterial cells for autoaggregation. In EHEC O26:H11, the expression of FdeC regulated by the histone-like nucleoid-structuring protein (H-NS) and occurring at temperatures above 39°C mediates biofilm formation and could contribute to colonization of the terminal rectum of cattle (Easton *et al.*, 2014).

7. Type VII secretion system (T7SS): Pili by numbers

In diderm-LPS bacteria, T7SS corresponds to the chaperone-usher (CU) pathway involved in the secretion and assembly of pili (Desvaux *et al.*, 2009). The nomenclature for pili formed via T7SS is messy, quite confusing, and essentially species-dependent; *e.g.* under the general denomination of Type 3 pili, the designations of coli surface (CS) pili, P pili, aggregative adherence fimbriae (AafD), adhesive fimbriae on RDEC-1 or diffuse adherence fibrillar adhesin/Dr blood group antigen (Afa/Dr) are also found, but all fall under the T7SS umbrella (Zav'yalov *et al.*, 2010; Busch and Waksman, 2012; Thanassi *et al.*, 2012). While numerous operons encoding such pili remain putative (Hayashi *et al.*, 2001; Low *et al.*, 2006b; Korea *et al.*, 2010), STEC secrete and assemble several pili via the T7SS. The operon encoding ECP (*E. coli* common pilus) is highly prevalent across *E. coli*, including EPEC (Saldana *et al.*, 2009). These pili participate in bacterial adhesion to human epithelial cells, intestinal colonization and biofilm development through interorganelle binding via the EcpA pilin (Rendon *et al.*, 2007). In *E. coli* O157:H7, these pili target arabinosyl residues in plant cell walls to mediate adhesion to vegetables (Rossez *et al.*, 2014). *E. coli* O157:H7 also contains two operons encoding the long polar fimbriae (Lpf), namely *lpf1* with six genes (*lpfABCC'DE*) forming LpfA pilus, and *lpf2* containing five genes (*lpfABCDD'*) forming LpfD2 pilus (Doughty *et al.*, 2002). Regarding the Lpf major subunits, five different genetic variants have been described for LpfA1 versus three for LpfA2 (Farfan and Torres, 2012). Lpf interact specifically with the

ECM, namely fibronectin, laminin or collagen IV components. They contribute to the adhesion of EHEC to the intestinal epithelium and can even influence bacterial tropism towards different intestinal tissues (Torres *et al.*, 2008; Farfan *et al.*, 2011). In EHEC O104:H4, LpfA contributes to bacterial adhesion to polarized intestinal epithelial cells, biofilm formation as well as cecum and large intestine colonization (Ross *et al.*, 2015). In EHEC, the *E. coli* YcbQ laminin-binding fimbriae (ELF) were found to bind laminin as well as to adhere to human epithelial cells (Samadder *et al.*, 2009a). The sorbitol-fermenting fimbriae (SFP) protein is responsible for hemagglutination activity in *E. coli* O157:H7, but also acts as adhesin via the SfpG pilin, which is involved in adherence to epithelial cells (Musken *et al.*, 2008). F9 pili are encoded in O-island 61 and are involved in the binding of EHEC to fibronectin and to bovine intestinal epithelial cells (Low *et al.*, 2006a). While encoded in numerous STEC (Li *et al.*, 1997), expression of Type 1 pili was only reported in a subset of STEC O26 and O118 strains (Enami *et al.*, 1999), where they participate in colonization of the bovine gut by mediating adhesion to intestinal epithelial cells, but also enable adhesion to abiotic surfaces (Galfi *et al.*, 1998; Cookson *et al.*, 2002). The expression of Type 1 pili is subject to phase variation of *fimA* encoding the major prepilin, but the environmental triggers remain to be determined (Li *et al.*, 1997; Roe *et al.*, 2001).

Like EAEC, EHEC O104:H4 lacks the LEE but encodes typical aggregative adhesion fimbriae (AAF) (Brzuszkiewicz *et al.*, 2011; Mellmann *et al.*, 2011). Just as all EPEC are AEEC, all EAEC are aggregative-adherence encoding *E. coli* (AAEC), *i.e. aaf*$^{+}$, from which atypical EHEC O104:H4 have emerged (Brzuszkiewicz *et al.*, 2011; Mellmann *et al.*, 2011; Tietze *et al.*, 2015). In EAEC, AAF are associated with a strong ability to form biofilms on biotic and abiotic surfaces as well as with hemagglutination with human erythrocytes [43]. Four variants of AAF have been identified, namely AAF/I to AAF/IV, encoded on virulence plasmids of the pAA family (Boisen *et al.*, 2008). AF/II binds to fibronectin (Farfan *et al.*, 2008). AAF act in concert with dispersin in the colonization of the intestinal mucosa, resulting in a highly virulent combination in EHEC O104:H4 when Stx is also present (Boisen *et al.*, 2015). While the expression of these different pili is subject to regulation by various environmental factors (Gonyar and Kendall, 2014), their global expression and respective contribution to the colonization process along the food chain, from the environment, animal reservoirs, food matrices to human infection, is far from understood. Besides, several operons encoding putative T7SS remain to be characterized in STEC.

8. Type VIII secretion system (T8SS): Curli

T8SS corresponds to the extracellular nucleation-precipitation (ENP) pathway involved in the secretion and assembly of peculiar pili, called curli (Desvaux *et al.*, 2009). In fact, curli are functional amyloid fibers predominantly composed of the major curli subunit protein CsgA following nucleation at the cell surface initiated by the minor curli subunit CsgB, which further promotes ramification along the fibers (Goyal *et al.*, 2014). Curli fibers are extremely adherent and are involved in cell aggregation, bacterial adhesion and, ultimately, biofilm development (Hammar *et al.*, 1996; Fronzes *et al.*, 2009). *E. coli* O157:H7 expressing curli are more virulent and exhibit greater adherence to eukaryotic cells than noncurliated strains (Uhlich *et al.*, 2002). The expression of curli seems to compensate for the absence of Lpf (Lloyd *et al.*, 2012). Rather than affecting initial attachment, curli enhance sessile development (Ryu and Beuchat, 2005).

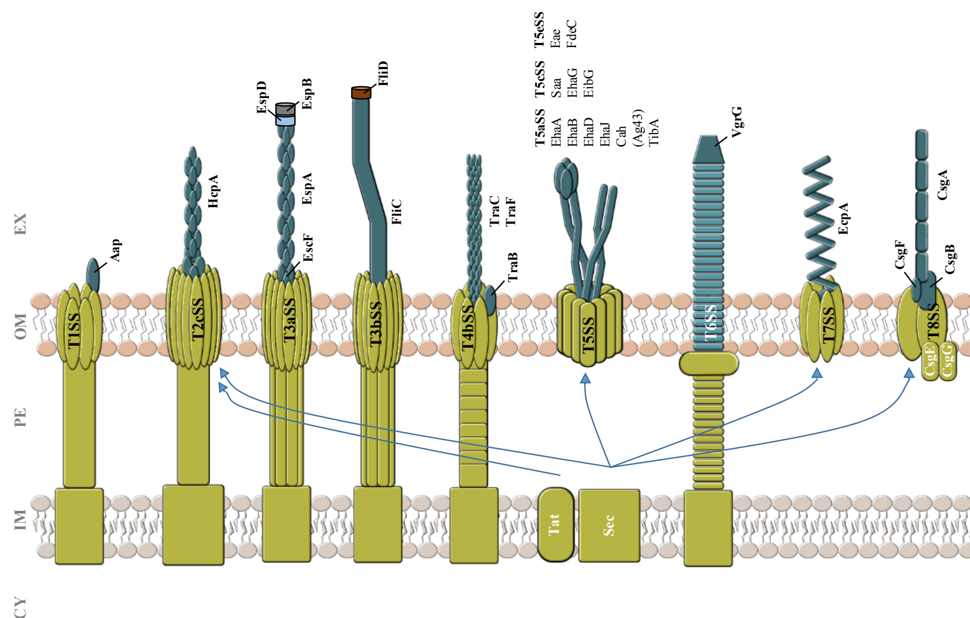


Figure 1: Overview of the complement of the secretome associated with surface colonization factors in STEC. Protein export systems are colored violet, protein secretion systems in diderm-LPS bacteria are colored green and their cognate effectors directly involved in surface colonization are colored blue. For details of the respective secretion systems and cognate colonization factors refer to the main text. T1SS: Type I protein secretion system or the ABC-MFP-TolC heterotrimeric system. T2SS: Type II protein secretion system or the secretion-dependent pathway (SDP), especially the type 4 pili (T4P) system (T2cSS). T3SS: Type III protein secretion system, including the injectisome (T3aSS) and flagellar (T3bSS) systems. T4SS: Type IV protein secretion system, especially the conjugative Tra system (T4bSS). T5SS: Type V protein secretion system, especially including the autotransporter (T5aSS), trimeric autotransporter (T5cSS) and inverted autotransporter (T5eSS) systems. T6SS: Type VI protein secretion system. T7SS: Type VII protein secretion system or the chaperone-usher (CU) pathway. T8SS: Type VIII protein secretion system or the extracellular nucleation-precipitation (ENP) pathway

Second part

Results

Chapter 1

The increasing incidence of EHEC infections, and of other diarrheal diseases caused by various etiological agents, is one of the reasons to pursue the development and discovery of new treatments to prevent these infections. In the case of EHEC, it is necessary to reduce the burden of the disease during the onset of infection since, paradoxically, the use of antibiotics is not recommended, and because of the emergence of antibiotic-resistant strains and collateral damage, such as hemorrhagic uremic syndrome (HUS). Increases in the number of cases of HUS have placed an increasing financial burden on healthcare systems, not only in developed countries, but also in countries in Latin America, as Argentina and Brazil. To date, there is no licensed human vaccine able to protect against enterohemorrhagic *E. coli*, either InPEC or ExPEC, and so there is a pressing need for a vaccine against pathogenic *E. coli*, and EHEC. As mentioned above, traditional approaches to vaccine development, ranging from the attenuation of pathogens to the identification of protective antigens or subunit vaccines, have been widely employed in the search for a suitable vaccine. However, traditional methods require years of work before a promising vaccine is finally produced. The new era of vaccine development has taken advantage of data generated by the "omics" technologies (genomics, proteomics, transcriptomics, etc.) to obtain new vaccines in a shorter period of time. Reverse vaccinology has become one of the best tools for vaccine development. It not only allows scrutiny of the entire coding DNA sequences of the genome, and prediction of their cellular localization and putative function, among other features, but also the combination of data, as from transcriptomics, the prediction of putative epitopes, and, most importantly, exploration of fastidious pathogens and how to get them to grow. The aim of the first part of this thesis consists of the identification of novel potential candidates against enterohemorrhagic *E. coli*, by a reverse vaccinology approach. During this work, bioinformatics approaches were applied to the EHEC O157:H7 EDL933 strain genome to predict the most promising vaccine candidates. Along with distribution analysis, we evaluated the prevalence and sequence variability of the selected antigens among other EHEC strains, serotypes and also other pathogenic *E. coli.* Subsequently, the antigens were obtained as recombinant proteins and expressed on generalized membrane modules for antigen delivery (GMMA), to evaluate their ability to confer immunoprotection and to reduce intestinal colonization in a mouse model.

Identification of a novel vaccine candidate against Enterohemorrhagic *E. coli* (EHEC) O157:H7 by Reverse Vaccinology approach

Maricarmen Rojas-Lopez[1,2], Manuele Martinelli[1], Valentina Brandi[3], Gregory Jubelin[2], Fabio Polticelli[3,4], Marco Soriani[1], Mariagrazia Pizza[1], Mickaël Desvaux[2], Roberto Rosini[1]*.

[1] GlaxoSmithKline Vaccines, Via Fiorentina 1, 53100, Siena, Italy.

[2] Université Clermont Auvergne, INRA, UMR454 MEDiS, 63000 Clermont-Ferrand, France.

[3] UniROMA Tre, Department of Sciences, Viale G. Marconi 446, 00146 Rome, Italy.

[4] National Institute of Nuclear Physics, Roma Tre Section, Rome, Italy.

*Corresponding author: Roberto Rosini, GSK, Via Fiorentina 1, 53100, Siena, Italy Tel.: +39 05 77 25 38 88, Fax: +39 05 77 25 38 88, e-mail: roberto.x.rosini@gsk.com

ABSTRACT

Enterohemorrhagic *E. coli* (EHEC) are a major cause of large outbreaks mainly affecting developed countries. From 1982 to 2002, a total of 350 *E. coli* O157 outbreaks were reported in the United States. EHEC infection causes diarrheal disease often associated with clinical complications like hemorrhagic colitis and hemolytic uremic syndrome (HUS).
Although efforts focused on hygiene have been implemented in the food supply chain to reduce the risk of the foodborne *E. coli* O157 infection, outbreaks caused by this pathogen are still common. In addition, antibiotic-based therapy is discouraged for their potential undesirable effect in releasing shiga-toxin from the bacteria. Among non-antibiotic preventing strategies, vaccine development is warranted, still nowadays a licensed vaccine specific for human use against EHEC is not available. In this study, we used the Reverse Vaccinology approach applied on the EHEC O157:H7 genome to select new potential vaccine candidates. We identified a panel of 24 of potential protein antigens and we successfully expressed three of them in Generalized Modules for Membrane Antigens (GMMA) delivery system. GMMA expressing these vaccine candidates resulted to be immunogenic, raising a specific antibody response for two of the selected antigens. In particular, immunization with MC001 candidate was able to reduce intestinal EHEC O157:H7 colonization lowering the bacterial count in feces, colon and ceacum tissues in mice. This candidate was found to be homologue to the *Salmonella* Typhimurium lipid A deacylase enzyme (LpxR) and to our knowledge this study was the first report describing it as vaccine candidate. Also, gene distribution and sequence variability analysis showed that MC001 was mainly present and conserved in EHEC O157:H7 and in some EPEC. Given the high genetic variability among and within these pathotypes, the identification and inclusion of this conserved candidate in a vaccine might cover against major intestinal pathogenic strains. This study provides a cost-effective approach of identifying not previously described immunogenic proteins to be used as potential vaccine candidates.

INTRODUCTION

Escherichia coli enterohemorrhagic (EHEC) is an anthropozoonose and etiology agent for diarrheal disease and hemorrhagic colitis. EHEC infections occur mainly in developed countries, the most often serotypes implicated in outbreaks are the O157:H7 serotype and the big six non-157 serotypes (O26:H11, O45:H2, O103:H2, O111:H8, O121:H19 and O145:H28) (Kaper et al., 2004; Moriel et al., 2012; Croxen et al., 2013). The main reservoir of EHEC is known to be ruminants and therefore the infection can mainly occur from direct or indirect

contamination of food products from animal feces (Rivas et al., 2016). The EHEC strains are characterized for the expression of the shiga-like toxin (Stx), being this the hallmark of the pathotype. Furthermore, some strains also carry the Locus of Enterocyte Effacement (LEE) that encodes the type three secretion system (T3SS) responsible for the attachment and effacing (A/E) on the intestinal microvilli (Kaper et al., 2004).

The implications after the infection with this pathogen, include the development of the hemolytic uremic syndrome (HUS) and later renal failure, due mainly by the Stx (Tarr et al., 2005). Although the use of antibiotic remains a key intervention strategy to treat many bacterial infections, this therapy is not recommended in diseases caused by EHEC (Goldwater and Bettelheim, 2012; Rivas et al., 2016). In fact, the use of antibiotics in the treatment against EHEC infections could lead to the cellular damage by increasing the production of stx toxin causing its release into the blood stream and further worsening the disease outcome (Pacheco and Sperandio, 2012).

Yet, the increasing burden of these *E. coli* diarrheal diseases, the emergence of hybrids strains (genomic plasticity) and the increasing annual cost for the health care systems point out the need to develop effective therapeutic and preventive strategies. Among these lasts, vaccination is the most promising strategy to control disease (Croxen and Finlay, 2010; Moriel et al., 2012; Croxen et al., 2013; Rappuoli et al., 2014)

So far, a number of vaccine candidates have been identified by different approaches. For instance, virulence factors expressed as recombinant proteins such as Shiga-toxin (Stx), Intimin and *E. coli* secreted protein A (EspA) or avirulent ghost cells of EHEC O157:H7 have been tested using different immunization routes and adjuvant combinations in a several animal models providing encouraged results (Rojas Lopez Submitted).

Additionally, previous studies have identified new promising vaccine candidates showing the potential of exploiting the Reverse Vaccinology approach (Pizza et al., 2000; Maione et al., 2005; Moriel et al., 2010; Garcia-Angulo et al., 2014). In particular, this strategy was performed on a complete sequence genome of a neonatal meningitis extraintestinal *E. coli* isolate (NMEC) leading to the identification of 230 potential antigens. Among these, the SslE protein, a conserved zinc metallopeptidase, was one of the most protective antigens by conferring protection in three different murine models (Moriel et al., 2010; Nesta et al., 2014). Moreover, a recent *in silico* approach aimed to develop DNA based vaccine identified new EHEC antigens, including among others a putative pilin subunit, T3SS structural protein (*escC*), and an outer membrane protein from the bacteriophage Bp933W (*iomW*) (Garcia-Angulo et al., 2014; Tapia et al., 2016)

In addition to the available technologies, new vaccine development strategies have been recently explored. These innovations ideally serve to make vaccine production simpler, more cost effective, improve antigen presentation and immune response (MacLennan and Saul, 2014). Outer membrane vesicles are one of these systems employed in the vaccine development. Gram-negative bacteria naturally release native outer membrane vesicles (NOMV) that are rich in outer membrane lipids, outer membrane and periplasmic proteins, presenting these antigens in their natural conformation (Ellis and Kuehn, 2010). Therefore, NOMV-based vaccines have been largely employed against the organism from which they are recovered (Nieves et al., 2014; Petersen et al., 2014) or to express and deliver heterologous antigens (Bartolini et al., 2013; Daleke-Schermerhorn et al., 2014; Fantappie et al., 2014). However, in native conditions NOMV are recovered in small quantity, thus *E. coli* strains can be genetically modified by deletion of the *tolR* gene to enhance the level of vesicles production (Bernadac et al., 1998). This system has been successfully used for expressing properly folded membrane associated recombinant antigens and to induce functional immune responses (Bartolini et al., 2013). Recently, this antigen delivery approach, also known as Generalized Modules for Membrane Antigens (GMMA), has been successfully implemented and largely used against different pathogens (Berlanda Scorza et al., 2012; Gerke et al., 2015; De Benedetto et al., 2017).

The main goal of our work was to identify novel antigens for vaccine development against infections caused by EHEC, using GMMA as delivery system. Our study led to the identification of a new potential vaccine candidate present in EHEC O157:H7 strains and able to reduce intestinal bacterial colonization level in mice.

MATERIAL AND METHODS

Bacterial strains and culture conditions

All bacterial strains were routinely grown in Luria Bertani (LB) media with antibiotic selection pressure if needed at 37ºC. The *E. coli* Mach1™-T1R (Thermofisher) strain was used for cloning while BL21-DE3 (NEB) strain was used to express and purify the antigens as recombinant proteins. EHEC O157:H7 strain EDL933 is the prototype pathotype used in this study for antigen identification by Reverse Vaccinology approach. The strain EHEC O157:H7 86-24 was used in the animal challenging experiments.

Vaccine candidate selection by Reverse Vaccinology

In silico vaccine candidate identification was performed as follows. The 5675 CDS of EHEC O157:H7 EDL933 strain annotated genome (GeneBank sequence CP008957.1) were

analyzed by PSORT software (Yu et al., 2010) to predict the subcellular localization. HTMM (http://www.cbs.dtu.dk/services/TMHMM (Krogh et al., 2001) was used for prediction of transmembrane regions in putative proteins. RNA-Seq mapping and RPKM (cutoff >10) calculation was performed using Geneious R9 software (http://www.geneious.com, (Kearse et al., 2012). Distribution and sequence variability analysis into 31 *E. coli* complete annotated genomes was performed by BLASTP (Altschul et al., 1997) using a cutoff of ≥90% of query coverage and a ≥80% of sequence protein identity. Only antigens present in more than 5 intestinal pathogenic *E. coli* strains were selected.

Cloning and recombinant protein production of vaccine candidates

All candidates were cloned and expressed as His-tagged fusion proteins without their predicted signal sequence. Prediction of the signal peptide was performed by Signal P (Nielsen, 2017). All fragments were amplified by PCR using primers enlisted in table S4 and using genomic DNA of *E. coli* EHEC O157:H7 EDL933 strain. The PCR amplicons were cloned into a pET-15b plasmid (Novagen, EMD Millipore) with a His-tag in the carboxyl-terminus by the polymerase incomplete primer extension (PIPE) method (Klock and Lesley, 2009) or the NEBuilder® HiFi DNA Assembly Master Mix (NEB). Plasmids were transformed in BL21-DE3 (NEB). Briefly, *E. coli* BL21-DE3 harboring pET-15b constructs were grown in EnPresso® (BioSilta) following the manufacturer suggestions and using IPTG 0.01M (Sigma-Aldrich) for induction. After biomass collection pellets were lysed by sonication. The suspension obtained was centrifuged and the supernatant passed through a Ni-NTA agarose chelating column (Qiagen). Proteins were eluted using imidazole-concentration gradient buffers. In the case of insoluble proteins, the immidazol-buffers for purification contained urea 6 M. Protein concentrations were measured by Pierce™ BCA Protein Assay Kit (Thermofisher).

Construction of TolR mutant

The *tolR* mutant in a *E. coli* K12 MC4100 strain was constructed by allelic maker exchange using the Lambda red system (Datsenko and Wanner, 2000). The *tolR* was interrupted with a chloramphenicol resistance cassette (*cat*). Briefly, the *cat* cassette was amplified using forward and reverse primers with ≈70-nucleotides tail homologous to the flanking region of *tolR*, (Table 2). The PCR product was purified and used to transform an *E. coli* K12 recipient cells (carrying the plasmid expressing the recombinase E, pKD46) as previously described (Datsenko and Wanner, 2000). The deletion of the *tolR* gene was confirmed by PCR genomic DNA amplification using primers specifically annealing to the genes upstream (*tolQ*) and

downstream (*tolA*) to *tolR*.

NOMV and GMMA production

For NOMV isolation, the *E. coli* K12 MC4100 WT strain was grown at 37°C in liquid LB medium. For GMMA production, *E. coli* K12 Δ*tolR*::*cat* was grown at 37°C in liquid LB medium containing chloramphenicol (20 µg/ml) as previously described (Fantappie et al., 2014; Rossi et al., 2016). Briefly, 75 ml of media were inoculated with *E. coli* K12 WT or Δ*tolR*::*cat* and grown at 37 °C, 150 rpm overnight (≈16 hrs.). To recover the supernatants cultures were centrifuged for 30 min at 8,000 x *g* and 0.22 µm filtered. These media were ultracentrifuged using propylene ultracentrifuge tubes (Beckman Coulter) at 105,000 x *g* for 2 hours at 4°C. Pellets were washed once with phosphate-buffered saline (PBS) and re-centrifuged. Finally, pellets were resuspended in 2 ml of PBS followed by 0.22-µm filtration and vesicles were stored at 4°C. To determine the total protein content present in these preparations a quantification was performed by DC protein assay (Bio-Rad) based on the Lowry assay (Rossi et al., 2015)

Negative-staining transmission electron microscopy

A drop of 10 µL of GMMA or NOMV suspension was placed on copper formvar/carbon-coated grids and adsorbed for 2 min. Grids were then washed with few drops of distilled water and blotted with a Whatman filter paper. For negative staining, grids were treated with Uranyless EM stain (Delta Microscopy with Chromalys, France) for 1 min, air-dried and viewed through transmission electron microscope Hitachi H-7650 at 80 kV. Electron micrographs were recorded at a nominal magnification of 120,000x.

Over-expression of antigens in GMMA

To overexpress the MC001, MC007 and MC020 candidates in GMMA, the corresponding coding sequences -including their own signal peptide- were cloned in frame into a pBAD-A plasmid (Thermofisher) using the NEBuilder® HiFi DNA Assembly Master Mix (NEB). Also, a FLAG-tag sequence (DYKDDDDK) was introduced between the signal peptide of each protein and the rest of their sequence. The generated constructs (pBAD-MC001F, pBAD-MC007F, pBAD-MC020F), were transformed into the *E. coli* K12 Δ*tolR*::*cat* mutant and induced with arabinose (0.01% final concentration). The purified GMMA were named GMMA-MC001, GMMA-MC007 and GMMA-MC020. GMMA not expressing any antigen and obtained by transforming the empty pBAD-A plasmid were named GMMA-K12.

Mice immunization and colonization model

Five weeks-old BALB/c mice (Janvier) (10 mice per group) were immunized with PBS-alum hydroxide as adjuvant (Alhydrogel® 2%, Invivogen), GMMA- expressing the vaccine candidates (GMMA-MC001, GMMA-MC007 or GMMA-MC020) plus adjuvant or with GMMA-K12 (GMMA not expressing any candidate). All GMMA preparations were formulated using 2 mg/ml alum hydroxide as adjuvant. Animals were immunized by intraperitoneal injections (i.p.) with 10 µg of GMMA plus adjuvant at day 1 and with 5 µg of GMMA plus adjuvant at day 21 and day 35. Blood was collected from all the mice prior immunization and two weeks after the third dose. The challenge experiment was performed 2 weeks after the last immunization and using the EHEC O157:H7 strain 86-24. Mice were treated with streptomycin 24-hours prior infection. Also, animal received cimetidine 2 hours before infection via i.p. The animals were infected with $5x10^9$ CFU via gavage. Animal monitoring was performed in a daily basis including weight, signs and symptoms surveillance. According to ethical rules, mice displaying signals of illness and losing more than 15% of the total weight were euthanized, collecting the colon and colon organs. Fecal pellets were collected every day from day 1 to day 7 post-infection. At day 7 the remaining mice were killed and organs were collected. This animal model was adapted from models previously reported (Mohawk and O'Brien, 2011; Garcia-Angulo et al., 2013; Garcia-Angulo et al., 2014). All animal experiments were reviewed and approved by the Auvergne Committee for Animal Experimentation C2EA (Agreement N°6065-2016071216144325V2).

Enzyme-linked Immunosorbent Assay (ELISA)

Ninety-six well Maxisorp plates (Nunc, Thermo Fisher Scientific) were coated with 1 µg/ml of GMMA preparations antigens or 1µg/ml of recombinant protein in PBS overnight (O/N) at 4 ºC. Next day, plates were washed 3 times with T-PBS (0.05% Tween 20 in PBS, pH 7.4) and blocked with 100 µl 2% BSA (Sigma® Aldrich) for 1 hour at 37 °C. Every incubation step was followed by triple T-PBS wash. Serum samples were initially diluted 1:200 in 2% BSA in T-PBS, transferred to coated-blocked plates and serially 2-fold diluted followed by 2-hours incubation at 37 ºC. Then 100 µl/well of 1:2,000 diluted alkaline phosphatase-conjugated goat anti-mouse IgG (H+L) (Southern Biotech) were added and incubated for 2 hours at 37 ºC. Bound alkaline phosphatase was visualized by adding SIGMAFAST p-Nitrophenyl phosphate (Sigma® Aldrich). After 30 minutes at room temperature, plates were analyzed at 405 nm in a microplate spectrophotometer. The endpoint titer of a sample is defined as the reciprocal of the

highest dilution that has a reading above the cut-off using the formula described by Frey and collaborators. (Frey et al., 1998).

Western Blotting

Western blots were carried out on whole cell extracts (wce), recombinant proteins or GMMA preparations. SDS page was performed in MES buffer (Thermofisher) and transferred to iBlot 2 nitrocellulose stacks (iBlot system, Thermofisher). To visualize transferred proteins, the membranes were stained with ponceau red. Then, membranes were blocked with 10% (w/v) blotting-grade blocker (Bio-Rad) in T-PBS. The membranes were later incubated with the respective mouse polyclonal antisera in a 1:1000 dilution in T-PBS-3% blocker 1 hr at room temperature. Membranes were washed three times with T-PBS and then incubated with goat anti-mouse horseradish peroxidase-conjugated IgG (Dako antibodies) diluted (1:2000) in T-PBS-3% blocker. Colorimetric staining was performed using Opti-4CN Substrate Kit (Bio-Rad) following manufacturer instructions. To detect the FLAG-tag monoclonal ANTI-FLAG® M2 secondary antibody was used (Sigma Aldrich).

Comparative structural modelling

Structural models of MC001 have been obtained by employing three different approaches: the threading/ab initio modelling method implemented in the I-TASSER pipeline (Roy et al., 2010), the membrane proteins-specific approach of MEMOIR (Ebejer et al., 2013) and the homology modelling method of SWISS-MODEL (Biasini et al., 2014). The search for suitable modelling templates has been carried out with PSI-BLAST (Position-Specific Iterated BLAST) (Altschul et al., 1997) sequence similarity search against the Protein Data Bank using the amino acid sequence of MC001 as a bait. While MEMOIR does not provide a proper quality assessment of the models, in the case of I-TASSER and SWISS-MODEL, the quality of the final models has been assessed through the parameters C-score and QMEAN4 (Benkert et al., 2009), respectively. The C-score is a confidence score calculated based on the reliability of threading template alignments and the convergence parameters of the structure assembly simulations. C-score values typically range between -5 and 2, higher values characterizing high confidence models and vice-versa. The QMEAN4 score is a linear combination of four statistical potential terms and is typically in the range 0-1, with higher values characterizing better quality models. MC001 models are characterized by a C-score of -5 and a QMEAN4 score of 0.74.

Statistical analysis

Statistical analysis was performed by using GraphPad Prism 7 software. Mann-Whitney (unpaired and non-parametric) and Student t tests with threshold of $P < 0.05$ were used to analyze the data of the bacterial counts from the mouse colonization model and for the IgG antibody response.

RESULTS

Identification of Vaccine candidates by Reverse Vaccinology

To identify potential antigens in EHEC O157:H7 EDL933 strain we applied the reverse vaccinology approach summarized in figure 1. The first step was to analyze the putative cellular localization of the 5675 coding sequences (CDS) from the annotated genome of EHEC O157:H7 EDL933 strain using the PSORT algorithm. We focused mainly in chromosome encoded proteins predicted to be exported, surface associated, outer membrane associated and with an unknown localization. The selection also included proteins greater than 200 amino acids and with less than 3 transmembrane repeats determined by the TMHMM algorithm. As a result of this analysis, we selected 329 potential vaccine candidates (Table 1S). Next, we exploited RNA-Seq data available in NCBI Sequence Read Archive (SRA) seeking for genes that were expressed at transcriptional level. These RNA-Seq dataset were generated using EHEC EDL933 strain grown in LB, LB with antibiotics, LB-agar media and cattle feces (Landstorfer et al., 2014). Reads mapping on EHEC EDL933 resulted in 68 genes showing an absolute index number of ≥ 10 RPKM (reads per kilobase per million mapped reads) in at least one of the four growth conditions analyzed (Table S2). Another selection criterion was based on gene variability and distribution analysis of these 68 EHEC EDL933 potential antigens on 31 complete genomes, to select those present (query coverage: ≥90%) and conserved (sequence identity ≥80%) in more than 5 different intestinal pathogenic *E. coli* strains (Figure 1 and Figure 1S). This *in silico* selection led to the identification of 24 potential antigens which were cloned, expressed and purified as recombinant His-tagged fusion proteins in *E. coli*. Of these, 12 were successfully purified as soluble and 12 as insoluble proteins (Table S3).

The recombinant proteins were then used to immunize mice in order to generate antigen-specific polyclonal antibodies. These antibodies were subsequently tested in Western Blot analysis to assess the expression level of the corresponding potential candidates in the homologous EHEC O157:H7 strain whole cell extract, leading to the identification of 17 expressed proteins in standard laboratory growth conditions (Table S3). As proof of concept

among these expressed proteins we selected three potential antigens satisfying all the criteria mentioned above to further investigate their antigenic potential. In particular, these three vaccine candidates included an outer membrane protein (MC001), a putative aminopeptidase (MC007) and an autotransporter belonging to the AIDA family (MC020) (Table 1).

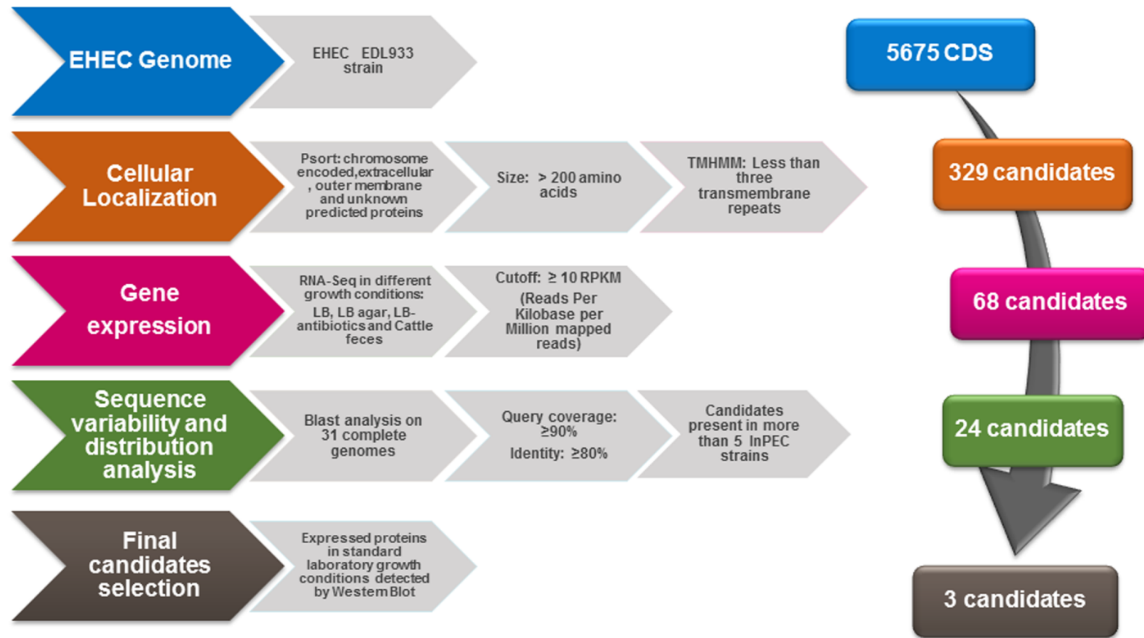


Figure 1. Vaccine candidate selection by Reverse Vaccinology in EHEC O157:H7. Schematic representation of the *in silico* pipeline used for vaccine candidate selection. The 5675 coding DNA sequences (CDS) chromosome-encoded of EHEC O157:H7 EDL933 strain were analyzed to identify potential antigens on the basis of their cellular localization (PSORT). The selection included proteins greater than 200 amino acids and with less than 3 transmembrane repeats determined by the TMHMM algorithm. RNA-Seq data were used to find candidates expressed at transcriptional level. Gene variability and distribution analysis of 62 potential antigens on 31 complete genomes were performed to select those present (query coverage: ≥90%) and conserved (sequence identity ≥80%) in more than 5 different intestinal pathogenic *E. coli* strains. Final selection was based also on those candidates expressed at the protein level on whole cell extracts.

Table 1. Features of the three selected candidates

Selected Candidates	Protein ID	Annotation feature	Localization prediction	Pfam Domain	Signal-Localization	Purified recombinant protein
MC001	AIG67060.1	Putative outer membrane protein	Extracellular	DUF2219	Non-Cytoplasmic	Insoluble
MC007	AIG66424.1	Putative aminopeptidase	Unknown	Unknown	Non-Cytoplasmic	Soluble
MC020	AIG69974.1	Pertactine precursor	Extracellular	AIDA, Pertactine	Unknown	Soluble

Antigen delivery in Generalized Modules for Membrane Antigens (GMMA)

To express the three selected candidates in GMMA, we first generated an overblebbing

E. coli K12 by mutating the *tolR* gene (K12 Δ*tolR*::*cat*). The native OMV (NOMV) released from the wild-type (wt) *E. coli* K12 and GMMA purified from the *tolR* mutant were observed by transmission electronic microscopy (TEM). This analysis showed that NOMV from wt strain appeared as closed spherical particles and homogeneous in shape with a size ranging from 20 to 100 nm (Figure 2A). On the other hand, GMMA from *tolR* mutant were released in higher amount (size ranging from 20 to 200 nm), and a small fraction of these vesicles showed an atypical shape characterized by more than one membrane layer (double-bilayer) a characteristic already observed by Perez-Crutz *et al.,* (Perez-Cruz et al., 2013; Perez-Cruz et al., 2016) (Figure 2B). Also, SDS-PAGE comparison of vesicles preparations showed that the *tolR* mutant was able to yield 25-fold more vesicles than wild-type strain in terms of total protein content, according to the amount of protein measured from GMMA preparations (Figure 2C). In order to express the three vaccine candidates (MC001, MC007 or MC0021) in GMMA, their coding sequence was cloned in frame onto pBAD plasmid and the FLAG-tag was inserted after their own signal peptide sequence (Figure 3A). The generated constructs (pBAD-MC001F, pBAD-MC007F, pBAD-MC020F) were transformed in the *tolR* mutant. To test whether these plasmids were expressing the vaccine candidates and incorporated into GMMA, we performed a western blot on the vesicles preparations using the anti-FLAG antibody. As shown in figure 3B all the three antigens were specifically recognized by the anti-FLAG antibody. These results indicate that the selected vaccine candidates were expressed in GMMA and these vesicles preparation can be used as antigen delivery system for animal model testing.

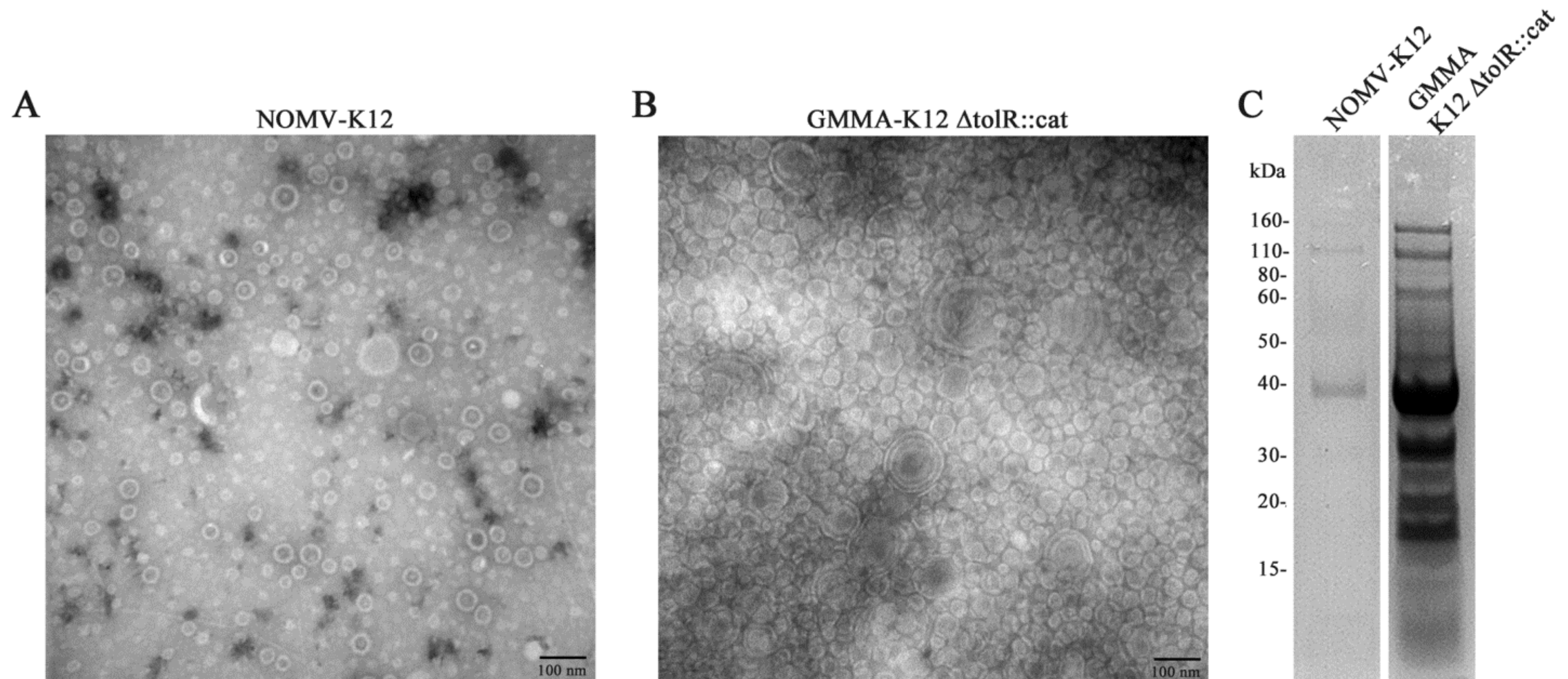


Figure 2. *E. coli* K12 engineering to generate Generalized Modules for Membrane Antigens (GMMA). NOMV and GMMA were isolated by ultracentrifugation from supernatants of *E. coli* K12 WT and K12 Δ*tolR::cat* (A) Negative staining of native NOMV released from a wild-type *E. coli* K12 observed by transmission electron microscopy (TEM). NOMV from K12 WT strain appeared as closed spherical particles and homogeneous in shape with a size ranging from 20 to 100 nm. (B) Negative staining of GMMA produced by the K12 *tolR::cat* (K12*tolR::cat*) strain analyzed by TEM GMMA from *tolR* mutant (size ranging from 20 to 200 nm) (magnification 120,000x). (C) SDS Page (4-12% bis-tris polyacrylamide) of membrane vesicles (NONV and GMMA) each purified from 75 ml of culture supernatants. Total protein content was quantified and 50 ug of GMMA obtained from K12 Δ*tolR::cat* sample was loaded into the SDS-PAGE gel. An equivalent volumetric amount of NOMV from K12 WT obtained from 75 mL of supernatant was loaded. The *tolR* mutant showed an extensive protein profile in the supernatant compared to wild type.

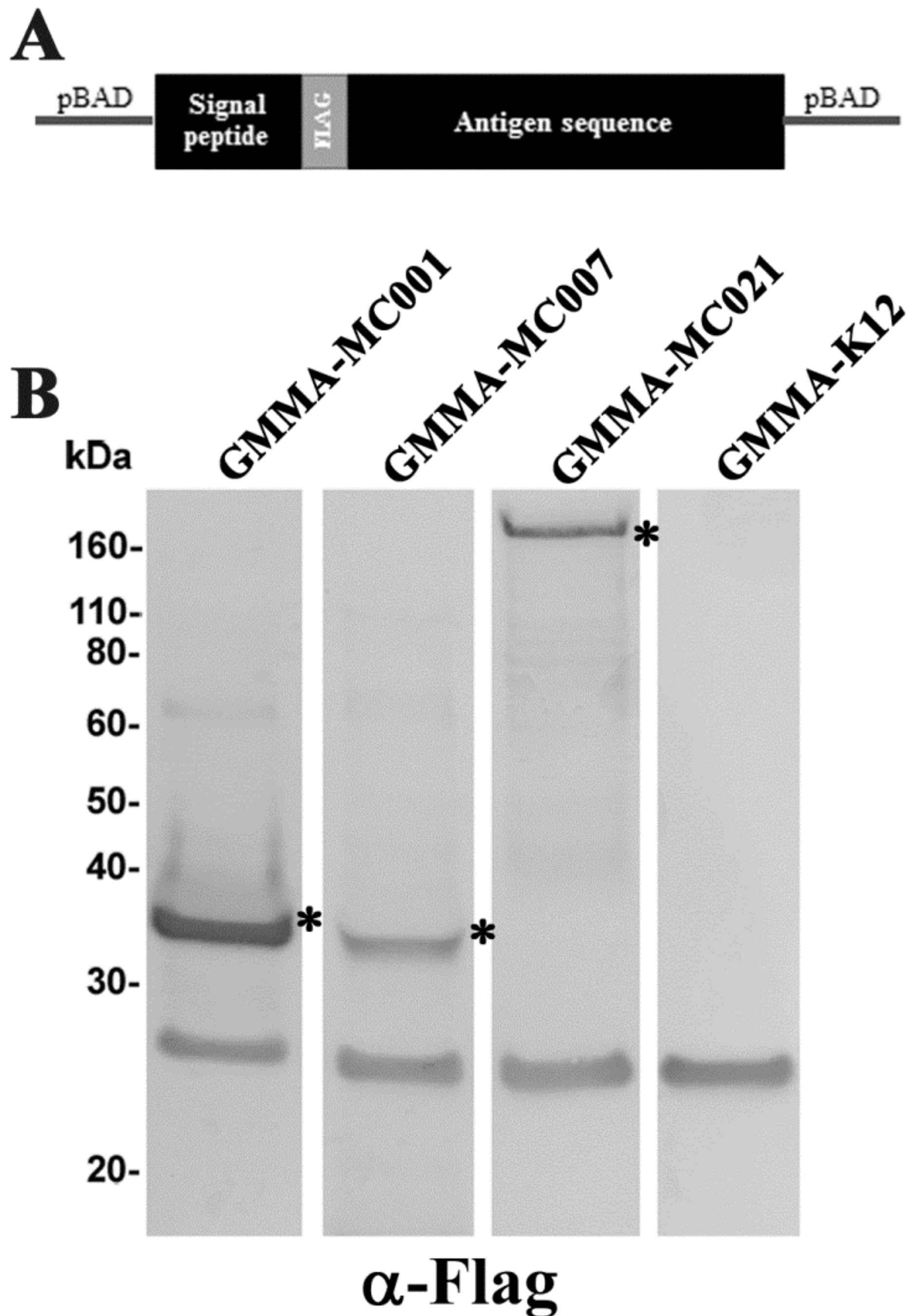


Figure 3. Antigen delivery into GMMA. (A) Schematic representation of the candidate cloning strategy. The three selected potential antigens were cloned into a pBAD vector using their own signal peptide. A FLAG-tag was inserted after the signal peptide of each construct. (B) Western blot of GMMA preparation expressing MC001, MC007 and MC021 candidates purified from the K12*tolR::cat* mutant using an anti-FLAG antibody. Asterisks (*) indicate the expected molecular size of each antigen.

Immunization with MC001-GMMA reduces EHEC intestinal bacterial colonization in mice

To test the possible ability of the selected potential antigens to prevent or reduce bacterial infection an intestinal colonization model was setup using BALB/c mice. Groups of ten mice were immunized with the GMMA vaccines over expressing the candidates or with empty GMMA-K12 via intraperitoneal at day 1, 21 and 39. At day 50, mice were infected with EHEC O157:H7 86-24 strain ($5x10^9$ CFU), via gavage. Fecal samples were collected in a daily basis performing bacterial counts. GMMA-MC001 immunized mice showed a ≈3-logs reduction (*P* value = 0.0001) in fecal bacteria number compared to PBS-alum at day 5 after infection while a ≈2-logs and ≈4-logs reduction (*P* value = 0.0033 and 0.0037) were obtained in comparison to empty GMMA-K12 immunized groups at day 6 and 7 respectively (Figure 4A). For ethical reasons, at day 5 most of the PBS-alum immunized mice were euthanized due to their weight loss (>15% of initial body weight). By contrast, immunized mice with GMMA-MC007 and GMMA-MC020 preparations did not show significant bacteria reduction in comparison to PBS-alum and GMMA-K12 immunized mice (Figure 4A). In addition, at day 7 post infection, colon and caecum tissues were collected from all mice groups and bacterial count was performed. The number of bacteria in colon and caecum tissues was significantly reduced (3-logs and 5-logs, *P* = 0.0003 and 0.0006 respectively) in mice immunized with the GMMA-MC001 in comparison to GMMA-K12 while GMMA-MC007 immunized mice showed a reduction of ≈2-logs and ≈3-logs for colon and caecum respectively (*P* = 0.0068 and 0.0012) (Figure 4B and 4C).

A

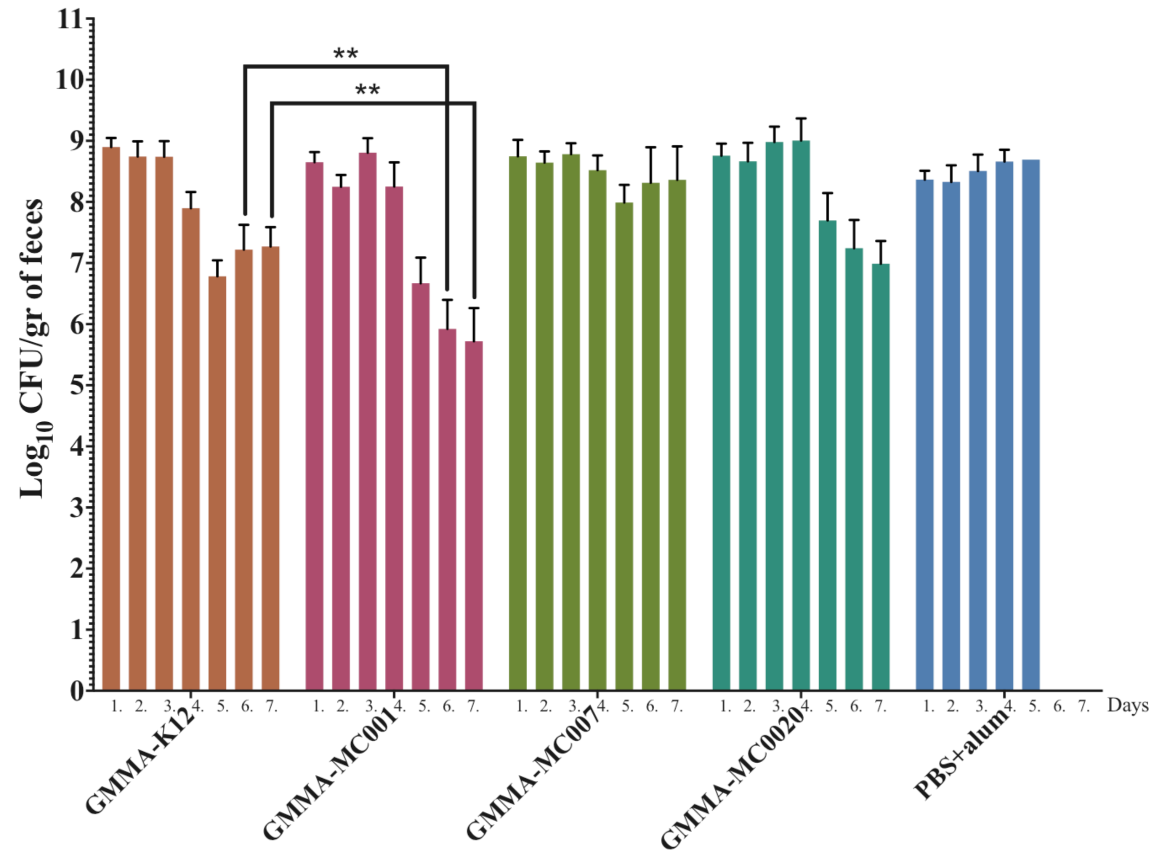
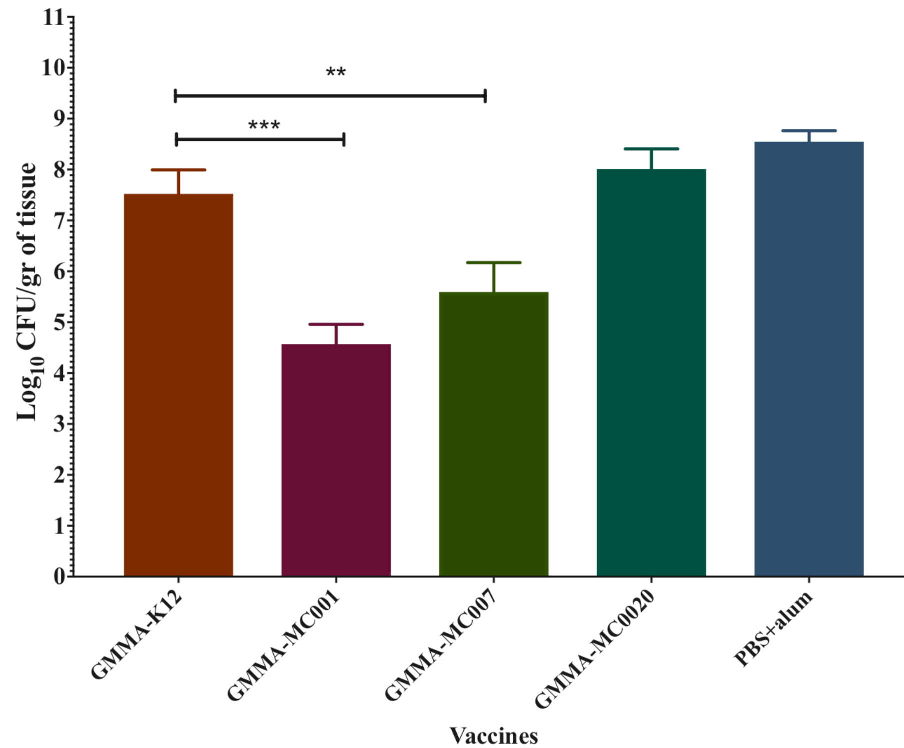
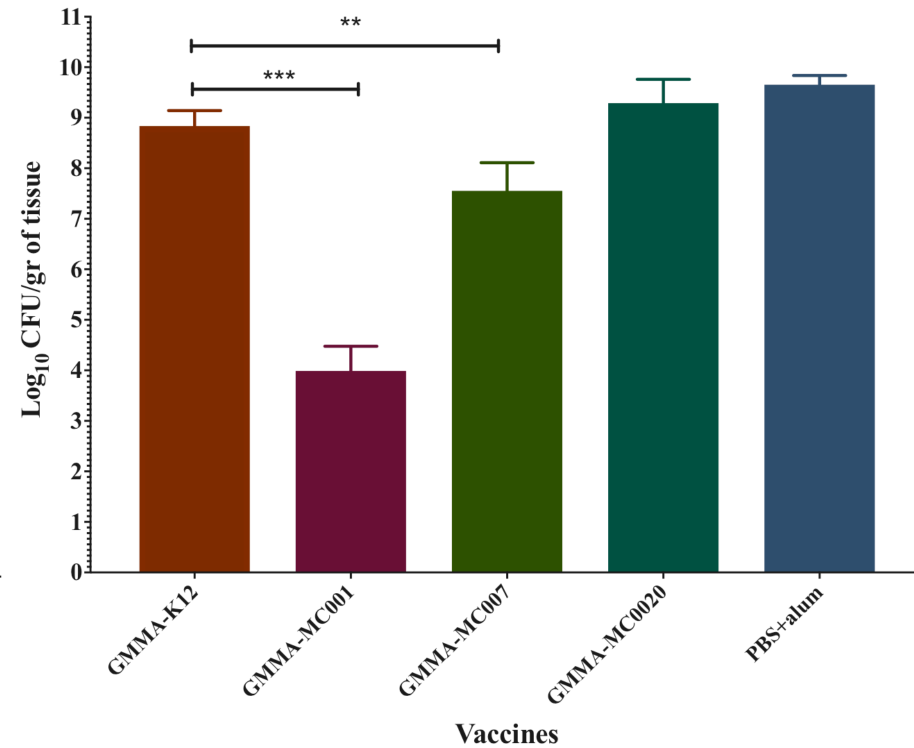


Figure 4. EHEC intestinal colonization model using mice immunized with GMMA overespressing vaccine candidates. (A) Graphic representation of 7-days bacterial counts in feces of immunized mice with GMMA vaccines (GMMA-K12, GMMA-MC001, GMMA-MC007, GMMA-MC020) overexpressing the candidates or with empty GMMA-K12 or PBS-alum via intraperitoneal at day 1, 21 and 39. Mice were challenged using EHEC O157:H7 86-24 strain via gavage. Fecal samples were collected in a daily basis performing bacterial counts expressed as CFU number per gr of feces (CFU/gr).

B



C

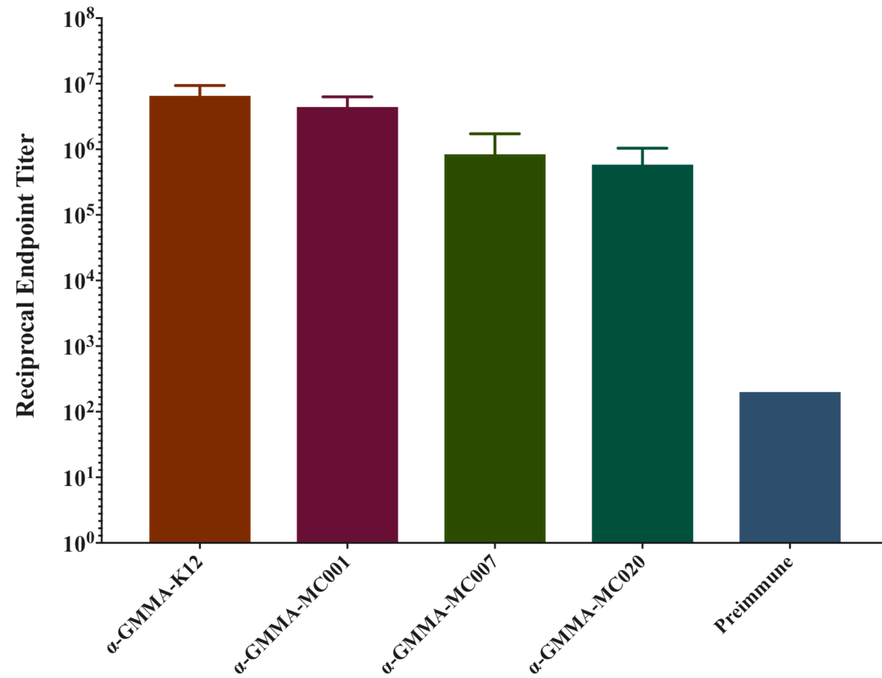


(B) Graphic representation of bacterial counts in colon tissue (CFU/gr) of immunized mice. (C) Graphic representation of bacterial counts in caecum tissue (CFU/gr) of immunized mice. The plots represent individual average. Data are expressed as means ± the standard deviation (SD) of values from ten mice in each group, and asterisks indicate statistically significant differences (*P* value < 0.05) (**, *P* value = 0.0033 and 0.0037; ***, *P* value = 0.0003 and 0.0006) calculated by Mann-Whitney test. Student t tests, indicates a significant difference in CFU number per gr of feces or tissue between mice vaccinated with GMMA-K12 (control group) and GMMA overexpressing the vaccine candidates.

Immunization with GMMA-expressing induces antigen specific antibody response for MC001 and MC021 vaccine candidates

To assess the immune response induced by GMMA immunization *per se* and the possible contribution in raising specific response against the three vaccine candidates (MC001, MC007 and MC020) overexpressed in GMMA their serum antibody level was determined by ELISA.

To measure the total level of immunoglobulins G (IgG), serum samples were collected prior the first shot of immunization (preimmune sera) and two weeks after the third immunization, before challenging mice. Microtiter plates coated with purified preparations of each GMMA-K12 and GMMA carrying the vaccine candidates showed higher total IgG levels in all the immunized groups versus the preimmune sera. Non-significant difference was found among the four GMMA immunized groups (Fig. 5A). On the other hand, to test whether there was an induction of a specific immune response attributable to the antigens expressed in GMMA, we performd ELISA assay using microtitre plates coated with the MC001, MC007 and MC020 recombinant proteins. A significant increase in antibody response was found for GMMA-MC001 (P = 0.0076) and GMMA-MC020 (P = 0.0075) sera in comparison to GMMA-K12 sera (Fig. 5B). To confirm that the immunization with GMMA expressing these vaccine candidates was able to generate antigen specific antibodies, we performed western blot assay, using MC001, MC020 or MC007 recombinant proteins as target. As shown in figure 5C, only MC001 and MC020 recombinant proteins were detected using GMMA-MC001 and GMMA-MC020 sera respectively, while MC007 was not recognized by GMMA-MC007 serum.

A 

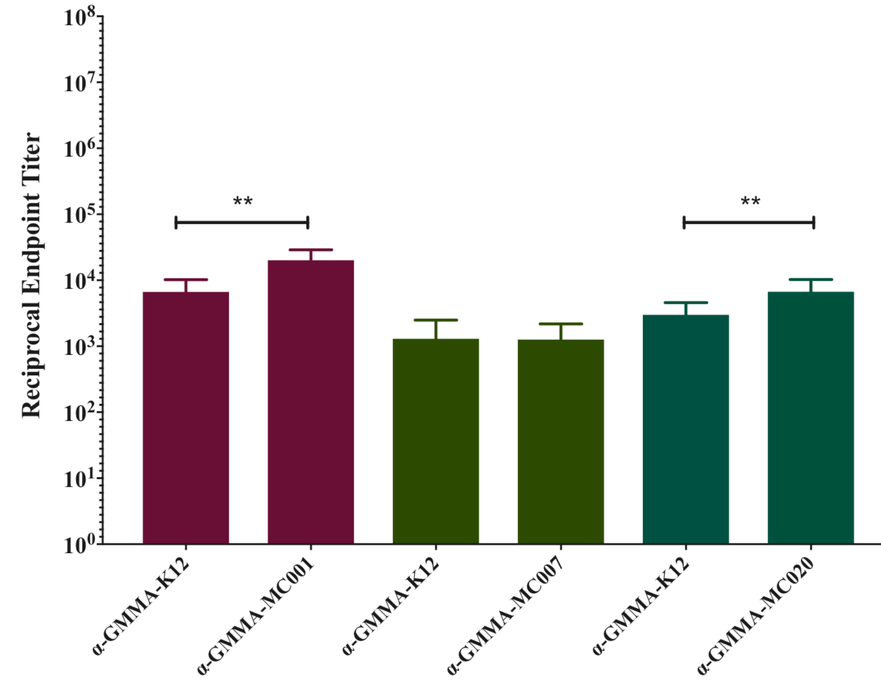
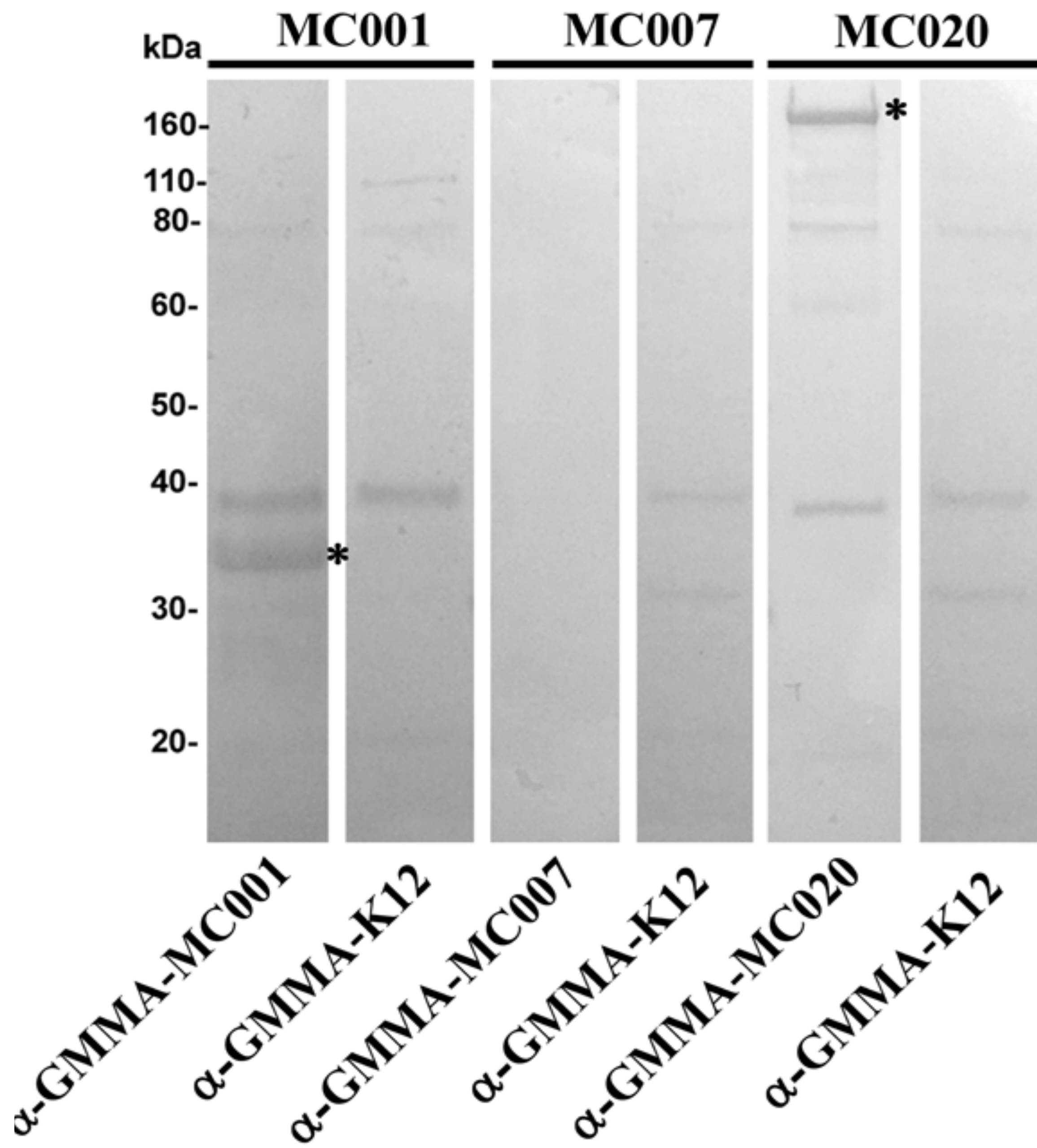
B 

Figure 5. Total IgG response after immunization and specific antibody detection of GMMA overexpressing vaccine candidates. (A) IgG antibody relative titers of GMMA overexpressing the antigens were measured by ELISA assay. Microtiter plates were coated with purified preparations of each GMMA-K12 and GMMA carrying the vaccine candidates. Sera raised against vaccine candidates overexpressed in GMMA were collected from vaccinated animals two weeks post the third immunization. Preimmune bar represents the sera pool of all groups. (B) Specific antibody (IgG) measurement of sera obtained by immunization of GMMA vaccines was performed by ELISA assay using microtitre plates coated with the MC001, MC007 and MC020 recombinant proteins. The plots represent individual average. Data are expressed as means ± the SD of values from ten mice in each group, and asterisks indicate statistically significant differences (P value < 0.05) (**, P value = 0.0076) calculated by Mann-Whitney test. Student t tests, indicates a significant difference in reciprocal endpoint titers between mice vaccinated with GMMA-K12 (control group) and GMMA overexpressing the vaccine candidates. The endpoint titer of a sample is defined as the reciprocal of the highest dilution that has a reading above the cutoff.

C



(C) Western blot assay, using MC001, MC020 or MC007 recombinant proteins as target. Each recombinant protein was tested using the respective serum raised against GMMA overexpressing the vaccine candidates and GMMA-K12 (negative control)

MC001 is homologous to *Salmonella* Typhimurium lipid A deacylase (LpxR)

To obtain more insights about the structural features of the MC001 vaccine candidate, an *in-silico* analysis was performed. In order to find proteins with known structure and significant sequence similarity with MC001, its protein sequence was used to run a PSI-BLAST search against the Protein Data Bank (PDB) (Berman et al., 2000). This search retrieved as first hit the sequence of the *Salmonella* Typhimurium lipid A deacylase (LpxR). Furthermore, a PSI-BLAST search over the non-redundant protein sequences database revealed a high sequence similarity also with LpxR from *Vibrio cholerae, Yersinia enterocolitica* and *Helicobacter pylori*. Structural MC001 models using the LpxR structure as a template (PDB code: 3FID) (Rutten et al., 2009) were built by I-TASSER, MEMOIR and SWISS-MODEL softwares. All the three generated models were in agreement with each other, showing a pairwise Cα root mean square deviation in all cases lower than 0.5 Å. A representative structural model obtained with the SWISS-MODEL is shown in Figure 6A. In particular, the MC001 model was composed by a 12-stranded β-barrel, in which the β-strands were arranged in an antiparallel fashion similarly to a structure that is quite common in porins and other cell membrane proteins. The high homology between MC001 and *Salmonella* Typhimurium LpxR was confirmed by the presence in the active site of six conserved residues essential for Ca+2 binding and LpxR catalytic activity: (*Salmonella*/*E. coli*) Asn (9/31), Asp (10/32), Thr/Ser (34/56), His (122/144), Gln (118/140) and Glu (128/150) (Figure 6B). Sequence comparison revealed that MC001 has the same length of the LpxR *Salmonella* orthologue (319 amino acids) and shares approximately 74% sequence identity and 93.73% sequence similarity (74% of identity with 93% query coverage) (Figure 6C).

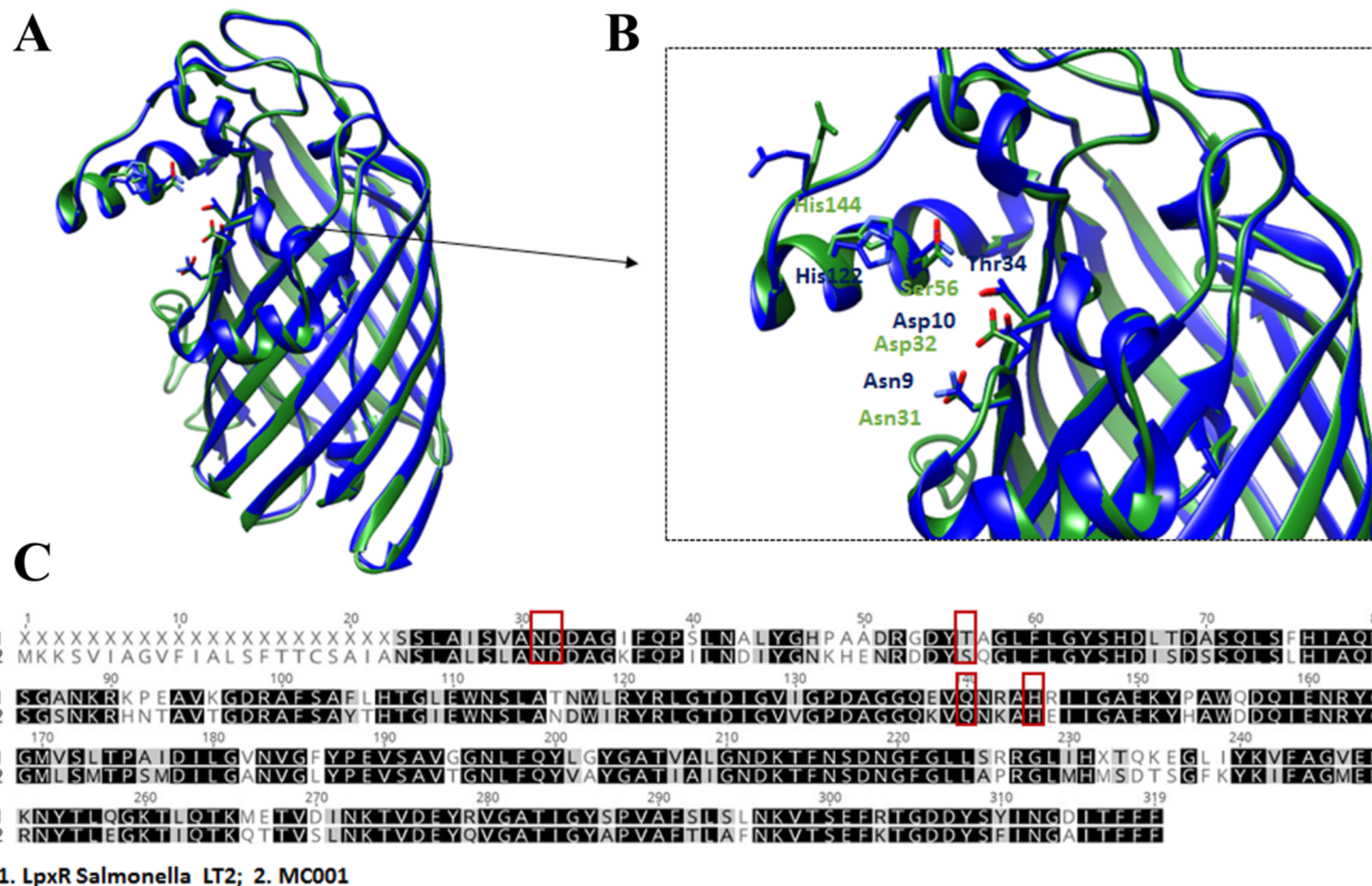


Figure 6. MC001 comparative protein modeling. (A) Structural model of MC001 obtained by SWISS-MODEL. Greens strips represent the 12-stranded β-barrel model for MC001 based on LpxR from *Salmonella* serovar Typhumurim (PDB: 3FID) model on blue strip (B) Magnification of the catalytic site region showing conserved residues for Ca+2 binding and LpxR catalytic activity: Asn (9/31), Asp (10/32), Thr/Ser (34/56), His (122/144), Gln (118/140) and Glu (128/150). (C) Sequence alignment of MC001 and LpxR shows an identity of 74% and a sequence similarity of 93% (black highlighted amino acids). The residues of the active site are marked in red squares.

DISCUSSION

In this study, we used the reverse vaccinology approach for the identification of new vaccine candidates in intestinal pathogenic *E. coli*, focusing on EHEC. In recent years, the reverse vaccinology has been exploited as an *in silico* tool to discover new protein antigens for several pathogens such as Serogroup B Meningococcus (MenB) (Pizza et al., 2000), Group B Streptococcus (Maione et al., 2005), extraintestinal *E. coli* (ExPECs) (Moriel et al., 2010) and for DNA vaccines against EHEC infections (Garcia-Angulo et al., 2014; Tapia et al., 2016). This approach allows the analysis of the whole annotated sequenced genome that includes all potential antigens (Rappuoli, 2000; Mora et al., 2003). In our study, we focus on chromosome encoded proteins predicted to be exported, surface associated, outer membrane associated and with an unknown localization. In fact, proteins which associate with bacterial membranes are more easily accessible to antibodies and therefore represent ideal vaccine candidates (Mora et al., 2006). As a result of the first step of this analysis, we identified 329 potential vaccine candidates starting from the 5675 CDS of the EHEC O157:H7 EDL933 annotated genome. Further, to determine whether these candidates are effectively expressed *in vitro* or *in vivo* during infection, with higher probability to be immunogenic, we took into account the expression at the RNA level. Exploiting previously generated RNA-Seq dataset (Landstorfer et al., 2014) we identified antigens expressed in at least one out of the four analyzed conditions, downselecting the potential candidates to a group of 68 putative antigens. An important characteristic for vaccine candidate is based on its ability to confer a broad-spectrum protection against the majority of circulating strains. However, given the high genetic variability among and within different *E. coli* pathotypes a universal vaccine candidate has not yet been identified. For instance, SslE a conserved and protective antigen identified in ExPEC is not present in EHEC pathotype (Moriel et al., 2010). Similarly, other proposed vaccine candidates, such as LEE and Stx, are not present in all the InPEC strains, nor even in some non-O157 EHEC strains, therefore limiting the vaccine use of these important virulence factors (Rojas-Lopez, submitted). For this reason, we searched for potential antigens possibly present and conserved in more than one single intestinal pathotype. Following these criteria, we selected 24 potential antigens. Moreover, 17 out of these 24 were also found to be expressed at the protein level in standard laboratory growth condition. Within this antigen panel, 12 proteins were predicted to be outer membrane associated proteins, three with extracellular localization and nine with unknown localization. On the basis of these observations we exploited the use of GMMA to express and deliver heterologous antigens in order to improve our chance to express properly folded membrane associated recombinant antigens and to induce functional immune responses,

as previously published (Berlanda Scorza et al., 2012; Bartolini et al., 2013; Daleke-Schermerhorn et al., 2014; Fantappie et al., 2014). GMMA can be also an ideal delivery system because of the presence of native proteins on the membrane surface that can act as potential self-adjuvants, helping to elicit immune responses (Kaparakis-Liaskos and Ferrero, 2015). However, in native conditions blebs are recovered in small quantity and, as consequence, *E. coli* strains need to be genetically modified by deleting the *tolR* gene to enhance the level of vesicles production (Bernadac et al., 1998; Berlanda Scorza et al., 2012). In a first attempt, we introduced the *tolR* mutation in a EHEC strain lacking also Stx to avoid possible release of toxins into the vesicles. However, once these overproduced blebs were used to immunize mice we observed sick animals, showing bristled hair, lethargic behavior and lose of weight (data not shown). For this reason, we constructed a *tolR* mutant in an avirulent *E. coli* K12 which was able to release higher amounts of blebs in comparison to the wild-type K12. In the present study, we used these GMMA-K12 for antigen expression and delivery. As proof of concept we selected three out of the 24 new promising antigens including an outer membrane protein (MC001), a putative aminopeptidase (MC007) and an autotransporter (MC020), that could have the potential to be expressed in GMMA. Two out of the three selected vaccine candidates (MC001 and MC020) showed typical features of membrane associated proteins, although the MC007 cellular localization was unknown, its signal sequence was predicted to direct the protein as a putative secreted protein. In addition, the MC001 candidate when purified as recombinant protein was obtained as insoluble form. From all these observations, the expression and delivery of these vaccine candidates in GMMA would increase their antigenic potential, with higher chance to be presented in their native conformation. Our data showed that all the three candidates were heterologous expressed into GMMA-K12 being specifically recognized by the anti-Flag antibody. For testing the possible ability of the selected candidates to prevent or reduce bacterial infection, an intestinal colonization model was setup using BALB/c mice. This animal model was adapted from the previous animal models of infection for EHEC and the one setup by Garcia-Angulo and colleagues (Mohawk and O'Brien, 2011; Garcia-Angulo et al., 2013). Although mice did not develop the symptoms associated with diarrheal disease as observed in humans, these murine models of *E. coli* O157:H7 infection, based on streptomycin-treated BALB/c mice, have been promising for EHEC colonization and candidate vaccine testing (Mohawk and O'Brien, 2011; Garcia-Angulo et al., 2013). Our data indicated that a stable EHEC intestinal bacterial colonization for 7-days post-infection using $5x10^9$ CFU was maintained and immunization with GMMA-K12 did not result toxic, as it was the case for GMMA-EHEC. We showed that intraperitoneal immunization of MC001 was able

to significantly reduce EHEC colonization in mice feces (day 6 and 7), colon and caecum tissues (day 7) in comparison to empty GMMA-K12. While for MC007 a less significant reduction was observed only in colon and caecum tissues. By contrast, we did not observe any EHEC colonization reduction for MC020 immunization.

Moreover, by assessing humoral response, higher specific titers of total IgG in GMMA-MC001 against the respective recombinant protein in the vaccinated group were observed. In support of these results western blot analysis showed that MC001 recombinant protein was recognized by the GMMA-MC001 serum indicating this vaccine candidate was associated with membrane vesicles and also easily accessible to the recognition by the immune system.

While MC020 recombinant protein was detected by its respective GMMA serum in western blot and higher IgG titers were sensed by ELISA and all in comparison to GMMA-K12, nonetheless this pattern was no observed for MC007. In fact, MC007 was only detected in GMMA preparation by the anti-FLAG antibody while serum raised against the GMMA expressing this antigen did not contained specific antibodies. Taken together these data might suggest that for a functional antigen delivery into the GMMA the heterologous antigen need to be expressed and properly presented to the immune system. As MC001 resulted in our study to be the most promising candidate, we used bioinformatic pipelines to predict molecular model by homology. BLAST analysis showed that MC001 was homologous to the *Salmonella* Typhimurium lipid A deacylase (LpxR) and shared similarity also with LpxR from *Vibrio cholerae, Yersinia enterocolitica* and *Helicobacter pylori*. Structural MC001 model using the *Salmonella* Typhimurium LpxR as a template revealed a structure composed by a 12-stranded β-barrel in which the β-strands that were arranged in an antiparallel fashion. The high homology between MC001 and *Salmonella* typhimurium LpxR was also confirmed by the presence in the active site of six conserved residues essential for Ca+2 binding and LpxR catalytic activity. In addition, gene variability and distribution analysis showed that MC001 was present and conserved among different EHEC strains as Sakai, O26:H11, O103:H2 and EPEC genomes.

Recently, it has been reported that LpxR can play an important role in pathogenesis by removing the 3'-acyloxyacyl group of lipid A (the hydrophobic anchor of lipopolysaccharide, LPS). This modification increases the ability of *Salmonella* Typhimurium to evade the innate immune response and promotes survival within macrophages (Kawasaki et al., 2012; Petrone et al., 2014). In this regard, targeting a specific antibody response toward LpxR could potentially avoid the LPS modification and subsequent immune evasion. However, further studies are ongoing to characterize this protein and its role in EHEC pathogenesis and immune

modulation in the host.

In conclusion, the reverse vaccinology approach applied on the EHEC O157:H7 genome combined with the GMMA antigen delivery system led us to identify new potential vaccine candidates. In particular, one of them (MC001) was able to reduce intestinal bacterial colonization and to our knowledge this study was the first report describing a lipid A deacylase enzyme (LpxR) as vaccine candidate. Given the high genetic variability among and within these pathotypes, the identification and inclusion of this conserved candidate in a vaccine might cover against major intestinal pathogenic strains. This study provided a cost-effective approach of identifying not previously described immunogenic proteins to be used as potential vaccine candidates.

ACKNOWLEDGEMENT

We thank you to the animal facility at INRA and Brigitte Gaillard-Martinie for transmission electron microscopy (TEM) at Université Clermont Auvergne.

AUTHOR CONTRIBUTIONS

MR-L, RR, MD, MS, MP conceived and designed the study. MR-L and RR performed the experiments. MR-L and MM designed constructs for recombinant protein expression and purification. MRL- and BG performed the TEM. MR-L, RR, MD, GJ, FP, MS and MP analyzed data. MR-L and RR, wrote the manuscript. All authors reviewed and approved the manuscript.

CONFLICT OF INTEREST STATEMENT

RR and MP are permanent employees of GSK. MM and MS were GSK employees at the time of the study. The authors declare that GSK provided support in the form of salaries. All the other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. Maricarmen ROJAS-LOPEZ is a Marie Curie PhD Research Fellow granted by ITN EID DISCo performed at GSK and INRA.

FUNDING

The study was funded by Marie Curie Actions DISCo program FP7-PEOPLE-2013-ITN (DISCo full partners: GSK/INRA, DISCo associated partners: ROMA TRE and Université Clermont-Auvergne).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found at the end of this thesis.

REFERENCES

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17)**,** 3389-3402.

Bartolini, E., Ianni, E., Frigimelica, E., Petracca, R., Galli, G., Berlanda Scorza, F., *et al.* (2013). Recombinant outer membrane vesicles carrying Chlamydia muridarum HtrA induce antibodies that neutralize chlamydial infection in vitro. *J Extracell Vesicles* 2. doi: 10.3402/jev.v2i0.20181.

Benkert, P., Kunzli, M., and Schwede, T. (2009). QMEAN server for protein model quality estimation. *Nucleic Acids Res* 37(Web Server issue)**,** W510-514. doi: 10.1093/nar/gkp322.

Berlanda Scorza, F., Colucci, A.M., Maggiore, L., Sanzone, S., Rossi, O., Ferlenghi, I., *et al.* (2012). High yield production process for Shigella outer membrane particles. *PLoS One* 7(6)**,** e35616. doi: 10.1371/journal.pone.0035616.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., *et al.* (2000). The Protein Data Bank. *Nucleic Acids Res* 28(1)**,** 235-242.

Bernadac, A., Gavioli, M., Lazzaroni, J.C., Raina, S., and Lloubes, R. (1998). Escherichia coli tol-pal mutants form outer membrane vesicles. *J Bacteriol* 180(18)**,** 4872-4878.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., *et al.* (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 42(Web Server issue)**,** W252-258. doi: 10.1093/nar/gku340.

Croxen, M.A., and Finlay, B.B. (2010). Molecular mechanisms of Escherichia coli pathogenicity. *Nat Rev Microbiol* 8(1)**,** 26-38. doi: 10.1038/nrmicro2265.

Croxen, M.A., Law, R.J., Scholz, R., Keeney, K.M., Wlodarska, M., and Finlay, B.B. (2013). Recent advances in understanding enteric pathogenic Escherichia coli. *Clin Microbiol Rev* 26(4)**,** 822-880. doi: 10.1128/CMR.00022-13.

Daleke-Schermerhorn, M.H., Felix, T., Soprova, Z., Ten Hagen-Jongman, C.M., Vikstrom, D., Majlessi, L., *et al.* (2014). Decoration of outer membrane vesicles with multiple antigens by using an autotransporter approach. *Appl Environ Microbiol* 80(18)**,** 5854-5865. doi: 10.1128/AEM.01941-14.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* 97(12)**,** 6640-6645. doi: 10.1073/pnas.120163297.

De Benedetto, G., Alfini, R., Cescutti, P., Caboni, M., Lanzilao, L., Necchi, F., *et al.* (2017). Characterization of O-antigen delivered by Generalized Modules for Membrane Antigens (GMMA) vaccine candidates against nontyphoidal Salmonella. *Vaccine* 35(3)**,** 419-426. doi: 10.1016/j.vaccine.2016.11.089.

Ebejer, J.P., Hill, J.R., Kelm, S., Shi, J., and Deane, C.M. (2013). Memoir: template-based structure prediction for membrane proteins. *Nucleic Acids Res* 41(Web Server issue)**,** W379-383. doi: 10.1093/nar/gkt331.

Ellis, T.N., and Kuehn, M.J. (2010). Virulence and immunomodulatory roles of bacterial outer membrane vesicles. *Microbiol Mol Biol Rev* 74(1)**,** 81-94. doi: 10.1128/MMBR.00031-09.

Fantappie, L., de Santis, M., Chiarot, E., Carboni, F., Bensi, G., Jousson, O., *et al.* (2014). Antibody-mediated immunity induced by engineered Escherichia coli OMVs carrying heterologous antigens in their lumen. *J Extracell Vesicles* 3. doi: 10.3402/jev.v3.24015.

Frey, A., Di Canzio, J., and Zurakowski, D. (1998). A statistically defined endpoint titer determination method for immunoassays. *J Immunol Methods* 221(1-2)**,** 35-41.

Garcia-Angulo, V.A., Kalita, A., Kalita, M., Lozano, L., and Torres, A.G. (2014). Comparative genomics and immunoinformatics approach for the identification of vaccine candidates for enterohemorrhagic Escherichia coli O157:H7. *Infect Immun* 82(5)**,** 2016-2026. doi: 10.1128/IAI.01437-13.

Garcia-Angulo, V.A., Kalita, A., and Torres, A.G. (2013). Advances in the development of enterohemorrhagic Escherichia coli vaccines using murine models of infection. *Vaccine* 31(32)**,** 3229-3235. doi: 10.1016/j.vaccine.2013.05.013.

Gerke, C., Colucci, A.M., Giannelli, C., Sanzone, S., Vitali, C.G., Sollai, L., *et al.* (2015). Production of a Shigella sonnei Vaccine Based on Generalized Modules for Membrane Antigens (GMMA), 1790GAHB. *PLoS One* 10(8)**,** e0134478. doi: 10.1371/journal.pone.0134478.

Goldwater, P.N., and Bettelheim, K.A. (2012). Treatment of enterohemorrhagic Escherichia coli (EHEC) infection and hemolytic uremic syndrome (HUS). *BMC Med* 10**,** 12. doi: 10.1186/1741-7015-10-12.

Kaparakis-Liaskos, M., and Ferrero, R.L. (2015). Immune modulation by bacterial outer membrane vesicles. *Nat Rev Immunol* 15(6)**,** 375-387. doi: 10.1038/nri3837.

Kaper, J.B., Nataro, J.P., and Mobley, H.L. (2004). Pathogenic Escherichia coli. *Nat Rev Microbiol* 2(2), 123-140. doi: 10.1038/nrmicro818.

Kawasaki, K., Teramoto, M., Tatsui, R., and Amamoto, S. (2012). Lipid A 3'-O-deacylation by Salmonella outer membrane enzyme LpxR modulates the ability of lipid A to stimulate Toll-like receptor 4. *Biochem Biophys Res Commun* 428(3), 343-347. doi: 10.1016/j.bbrc.2012.10.054.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., *et al.* (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12), 1647-1649. doi: 10.1093/bioinformatics/bts199.

Klock, H.E., and Lesley, S.A. (2009). The Polymerase Incomplete Primer Extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis. *Methods Mol Biol* 498, 91-103. doi: 10.1007/978-1-59745-196-3_6.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305(3), 567-580. doi: 10.1006/jmbi.2000.4315.

Landstorfer, R., Simon, S., Schober, S., Keim, D., Scherer, S., and Neuhaus, K. (2014). Comparison of strand-specific transcriptomes of enterohemorrhagic Escherichia coli O157:H7 EDL933 (EHEC) under eleven different environmental conditions including radish sprouts and cattle feces. *BMC Genomics* 15, 353. doi: 10.1186/1471-2164-15-353.

MacLennan, C.A., and Saul, A. (2014). Vaccines against poverty. *Proc Natl Acad Sci U S A* 111(34), 12307-12312. doi: 10.1073/pnas.1400473111.

Maione, D., Margarit, I., Rinaudo, C.D., Masignani, V., Mora, M., Scarselli, M., *et al.* (2005). Identification of a universal Group B streptococcus vaccine by multiple genome screen. *Science* 309(5731), 148-150. doi: 10.1126/science.1109869.

Mohawk, K.L., and O'Brien, A.D. (2011). Mouse models of Escherichia coli O157:H7 infection and shiga toxin injection. *J Biomed Biotechnol* 2011, 258185. doi: 10.1155/2011/258185.

Mora, M., Donati, C., Medini, D., Covacci, A., and Rappuoli, R. (2006). Microbial genomes and vaccine design: refinements to the classical reverse vaccinology approach. *Curr Opin Microbiol* 9(5), 532-536. doi: 10.1016/j.mib.2006.07.003.

Mora, M., Veggi, D., Santini, L., Pizza, M., and Rappuoli, R. (2003). Reverse vaccinology. *Drug Discov Today* 8(10), 459-464.

Moriel, D.G., Bertoldi, I., Spagnuolo, A., Marchi, S., Rosini, R., Nesta, B., *et al.* (2010). Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic Escherichia coli. *Proc Natl Acad Sci U S A* 107(20), 9072-9077. doi: 10.1073/pnas.0915077107.

Moriel, D.G., Rosini, R., Seib, K.L., Serino, L., Pizza, M., and Rappuoli, R. (2012). Escherichia coli: great diversity around a common core. *MBio* 3(3). doi: 10.1128/mBio.00118-12.

Nesta, B., Valeri, M., Spagnuolo, A., Rosini, R., Mora, M., Donato, P., *et al.* (2014). SslE elicits functional antibodies that impair in vitro mucinase activity and in vivo colonization by both intestinal and extraintestinal Escherichia coli strains. *PLoS Pathog* 10(5), e1004124. doi: 10.1371/journal.ppat.1004124.

Nielsen, H. (2017). Predicting Secretory Proteins with SignalP. *Methods Mol Biol* 1611, 59-73. doi: 10.1007/978-1-4939-7015-5_6.

Nieves, W., Petersen, H., Judy, B.M., Blumentritt, C.A., Russell-Lodrigue, K., Roy, C.J., *et al.* (2014). A Burkholderia pseudomallei outer membrane vesicle vaccine provides protection against lethal sepsis. *Clin Vaccine Immunol* 21(5), 747-754. doi: 10.1128/CVI.00119-14.

Perez-Cruz, C., Canas, M.A., Gimenez, R., Badia, J., Mercade, E., Baldoma, L., *et al.* (2016). Membrane Vesicles Released by a hypervesiculating Escherichia coli Nissle 1917 tolR Mutant Are Highly Heterogeneous and Show Reduced Capacity for Epithelial Cell Interaction and Entry. *PLoS One* 11(12), e0169186. doi: 10.1371/journal.pone.0169186.

Perez-Cruz, C., Carrion, O., Delgado, L., Martinez, G., Lopez-Iglesias, C., and Mercade, E. (2013). New type of outer membrane vesicle produced by the Gram-negative bacterium Shewanella vesiculosa M7T: implications for DNA content. *Appl Environ Microbiol* 79(6), 1874-1881. doi: 10.1128/AEM.03657-12.

Petersen, H., Nieves, W., Russell-Lodrigue, K., Roy, C.J., and Morici, L.A. (2014). Evaluation of a Burkholderia pseudomallei Outer Membrane Vesicle Vaccine in Nonhuman Primates. *Procedia Vaccinol* 8, 38-42. doi: 10.1016/j.provac.2014.07.007.

Petrone, B.L., Stringer, A.M., and Wade, J.T. (2014). Identification of HilD-regulated genes in Salmonella enterica serovar Typhimurium. *J Bacteriol* 196(5), 1094-1101. doi: 10.1128/JB.01449-13.

Pizza, M., Scarlato, V., Masignani, V., Giuliani, M.M., Arico, B., Comanducci, M., *et al.* (2000). Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 287(5459), 1816-1820.

Rappuoli, R. (2000). Reverse vaccinology. *Curr Opin Microbiol* 3(5), 445-450.

Rappuoli, R., Pizza, M., Del Giudice, G., and De Gregorio, E. (2014). Vaccines, new opportunities for a new society. *Proc Natl Acad Sci U S A* 111(34)**,** 12288-12293. doi: 10.1073/pnas.1402981111.

Rivas, M., Chinen, I., and Guth, B.E.C. (2016). "Enterohemorrhagic (Shiga Toxin-Producing) Escherichia coli," in *Escherichia coli in the Americas,* ed. A.G. Torres. (Cham: Springer International Publishing), 97-123.

Rossi, O., Caboni, M., Negrea, A., Necchi, F., Alfini, R., Micoli, F., *et al.* (2016). Toll-Like Receptor Activation by Generalized Modules for Membrane Antigens from Lipid A Mutants of Salmonella enterica Serovars Typhimurium and Enteritidis. *Clin Vaccine Immunol* 23(4)**,** 304-314. doi: 10.1128/CVI.00023-16.

Rossi, O., Maggiore, L., Necchi, F., Koeberling, O., MacLennan, C.A., Saul, A., *et al.* (2015). Comparison of Colorimetric Assays with Quantitative Amino Acid Analysis for Protein Quantification of Generalized Modules for Membrane Antigens (GMMA). *Molecular Biotechnology* 57(1)**,** 84-93. doi: 10.1007/s12033-014-9804-7.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5(4)**,** 725-738. doi: 10.1038/nprot.2010.5.

Rutten, L., Mannie, J.P., Stead, C.M., Raetz, C.R., Reynolds, C.M., Bonvin, A.M., *et al.* (2009). Active-site architecture and catalytic mechanism of the lipid A deacylase LpxR of Salmonella typhimurium. *Proc Natl Acad Sci U S A* 106(6)**,** 1960-1964. doi: 10.1073/pnas.0813064106.

Tapia, D., Ross, B.N., Kalita, A., Kalita, M., Hatcher, C.L., Muruato, L.A., *et al.* (2016). From In silico Protein Epitope Density Prediction to Testing Escherichia coli O157:H7 Vaccine Candidates in a Murine Model of Colonization. *Front Cell Infect Microbiol* 6**,** 94. doi: 10.3389/fcimb.2016.00094.

Tarr, P.I., Gordon, C.A., and Chandler, W.L. (2005). Shiga-toxin-producing Escherichia coli and haemolytic uraemic syndrome. *Lancet* 365(9464)**,** 1073-1086. doi: 10.1016/S0140-6736(05)71144-2.

Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., *et al.* (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26(13)**,** 1608-1615. doi: 10.1093/bioinformatics/btq249.

Table 4 supplementary material. Primers used in this study

Primer	Sequence	Product	Notes (ID, purpose)
MCRL-34Fw	CTGTACTTCCAGGGCTTTCAACCAATACTTAATGAT	MC001	AIG67060.1
MCRL-34Rv	AATTAAGTCGCGTTAAAAAAAGAAGGTGATTGC		
MCRL-24Fw	CTGTACTTCCAGGGCGATATCAATCTGTATGGTCC	MC002	AIG68165.1
MCRL-24Rv	AATTAAGTCGCGTTAAGTGCCTTTCCTGGT		
MCRL-40Fw	CTGTACTTCCAGGGCGACCTTTATGGCAAGG	MC003	AIG68357.1
MCRL-40Rv	AATTAAGTCGCGTTAGAACTGATAGGTAATGCC		
MCRL-35Fw	CTGTACTTCCAGGGCGAGTTTACGATAGACTTTTCG	MC004	AIG67577.1
MCRL-35Rv	AATTAAGTCGCGTTATTTACCCGTTGTATATAAAAAC		
MCRL-32Fw	CTGTACTTCCAGGGCACGTATGTAGATTCGCTG	MC005	AIG69411.1
MCRL-32Rv	AATTAAGTCGCGTTAACTGCTAATAGTTCTGCG		
MCRL-41Fw	CTGTACTTCCAGGGCATGAAGATAAGCTCTACT	MC006	AIG66347.1
MCRL-41Rv	AATTAAGTCGCGTTATTCGTAGGTAAAGGA		
MCRL-28Fw	CTGTACTTCCAGGGCCAGCAACTGACAGACAA	MC007	AIG66424.1
MCRL-28Rv	AATTAAGTCGCGTTACTGGAATCGACTCACC		
MCRL-36Fw	CTGTACTTCCAGGGCAACTGCTATTTTGGTACC	MC008	AIG66984.1
MCRL-36Rv	AATTAAGTCGCGTTAGTTGTAGTTAATTTTGAAAAG		
MCRL-38Fw	CTGTACTTCCAGGGCATGTGGGAATGTGATG	MC009	AIG67652.1
MCRL-	AATTAAGTCGCGTTATTGCATTTTCACCAA		

38Rv			
MCRL-26Fw	CTGTACTTCCAGGGCACGCTGACGGTAACGG	MC010	AIG67671.1
MCRL-26Rv	AATTAAGTCGCGTTATTTCGCACCTCGCTG		
MCRL-43Fw	CTGTACTTCCAGGGCACGCTTTACGAGCAG	MC011	AIG67672.1
MCRL-43Rv	AATTAAGTCGCGTTATAGATAGTATTTAAAGCCGT		
MCRL-22Fw	CTGTACTTCCAGGGCGTGGGGATCGACAGC	MC012	AIG68053.1
MCRL-22Rv	AATTAAGTCGCGTTAGAACGACCAGTTCACAC		
MCRL-42Fw	CTGTACTTCCAGGGCATGAAATTCCCTTCA	MC013	AIG68360.1
MCRL-42Rv	AATTAAGTCGCGTTAGTGATAAAAAGGCCA		
MCRL-23Fw	CTGTACTTCCAGGGCGGTTATTTTGTTGGCG	MC014	AIG68899.1
MCRL-23Rv	AATTAAGTCGCGTTAATCCTGCAACGCATA		
MCRL-29Fw	CTGTACTTCCAGGGCATTTATTATGCGATGAAA	MC015	AIG69664.1
MCRL-29Rv	AATTAAGTCGCGTTACAGTTCATTCAGTACATACTG		
MCRL-30Fw	CTGTACTTCCAGGGCATGGGTACAGCAGCTATA	MC016	AIG70798.1
MCRL-30Rv	AATTAAGTCGCGTTATTGTATTCCTGCACCA		
MCRL-37Fw	CTGTACTTCCAGGGCGATGTCCGGCGTAGC	MC017	AIG67308.1
MCRL-37Rv	AATTAAGTCGCGTTACTGCTTTTTACAACCATTC		
MCRL-48Fw	CTGTACTTCCAGGGCATGAGCGGAAAACCG	MC018	AIG66972.1
MCRL-48Rv	AATTAAGTCGCGTTATTTTCTTATTCCTCTCGATG		
MCRL-52Fw	CTGTACTTCCAGGGCATGGCTGCGGCAGCA	MC019	AIG69216.1
MCRL-	AATTAAGTCGCGTTACGCAATATTGACGAT		

52Rv			
MCRL-49Fw	CTGTACTTCCAGGGCGTGGGGCAGTCTAAT	MC020	AIG69974.1
MCRL-49Rv	AATTAAGTCGCGTTAACTCGCTTTCATTATGTT		
MCRL-51Fw	CTGTACTTCCAGGGCGTGCCTTACACGCTTGGT	MC021	AIG71811.1
MCRL-51Rv	AATTAAGTCGCGTTAAAGTGATTTACGGCAGGC		
MCRL-45Fw	CTGTACTTCCAGGGCATGGTCGCTAAATTAAAAC	MC022	AIG71181.1
MCRL-45Rv	AATTAAGTCGCGTTAAGCCTGGGTTATATTAAC		
MCRL-45Fw	CTGTACTTCCAGGGCATGGTCGCTAAATTAAAAC	MC023	AIG66227.1
MCRL-45Rv	AATTAAGTCGCGTTAAGCCTGGGTTATATTAAC		
MCRL-46Fw	CTGTACTTCCAGGGCAGTTATGGCCGATTT	MC024	AIG66656.1
MCRL-46Rv	AATTAAGTCGCGTTATGTAAACTGCACATAAGA		
pET-TEV-Fw	TAACGCGACTTAATTCTAGCATAACC	pET15	pET-15 expression vector
pET-TEV-Rv	GCCCTGGAAGTACAGGTTTTC		
TolR-Cat-Fw	TGGAGGTCGATTTGCCAGACGCTACTGAATCACAGGCGGTGAGCAGTAACGATAATCCGCCAGTGATTGTTGACATATGAATATCCTCCTTAGTTCCTATTCC	Cat-TolR	*tolR* mutant construction
TolR-Cat-Rv	CCTTGAAACGGCTGGACACTTCCGCCACCACCTGCTCTGGTGGTAAACGCTCCAGGCGATCTTTCTCAACCAGTGTAGGCTGGAGCTGCTTCGAAG		
GMMA-MC001Fw	CAGGAGGAATTAACCATGAAAAAAAGTGTCATCGCTGGC	MC001 with signal sequence	Antigen expression on GMMA
GMMA-MC001Rv-NONhis	ATGATGATGATGATGATGAAAAAAGAAGGTGATTGCTCC		
GMMA-MC007Fw	CAGGAGGAATTAACCATGGGGCAGTCTAATAATACCAC	MC007 with signal	Antigen expression

GMMA-MC007Rv-NONhis	ATGATGATGATGATGATGGAACGACCAGTTCACACCAGC	sequence	on GMMA
GMMA-MC020Fw	CAGGAGGAATTAACCATGGGGCAGTCTAATAATACCACC	MC020 with signal sequence	Antigen expression on GMMA
GMMA-MC020Rv-NONhis	ATGATGATGATGATGATGGAACGACCAGTTCACACCAG		
MC001tag Fw	GATTACAAAGACGATGATGACAAGGATTACAAAGACGATGATGACAAGAACAGCCTTGCATTATCATT	MC001 with signal sequence and Flag Tag	Antigen expression on GMMA with Flag Tag
MC001tag Rv	CTTGTCATCATCGTCTTTGTAATCCTTGTCATCATCGTCTTTGTAATCCGCGATAGCTGAACACGTGG		
MC007tag Fw	GATTACAAAGACGATGATGACAAGGATTACAAAGACGATGATGACAAGTCTGAATCTTCTATTGATGA	MC007 with signal sequence and Flag Tag	Antigen expression on GMMA with Flag Tag
MC007tag Rv	CTTGTCATCATCGTCTTTGTAATCCTTGTCATCATCGTCTTTGTAATCGGCACATCCTGCAAGCAGCG		
MC020tag Fw	GATTACAAAGACGATGATGACAAGGATTACAAAGACGATGATGACAAGGCCTTCACTCCTGATGTTAT	MC020 with signal sequence and Flag Tag	Antigen expression on GMMA with Flag Tag
MC020tag Rv	CTTGTCATCATCGTCTTTGTAATCCTTGTCATCATCGTCTTTGTAATCAGCCATCCCGGGCGGGGCAT		
pBADRv	CATGGTTAATTCCTCCTGTTAGCCC	pBAD Vector without His-Tag	Expression of Ag in GMMA
pBADFw NonHIS	TGAGTTTAAACGGTCTCCAGCTTGG		

Vaccine candidate name	Intestinal *Escherichia coli*																	Extraintestinal *Escherichia coli*													
	EHEC							EPEC		ETEC			AIEC			EAEC		NMEC			UPEC								AREC	APEC	ABU
	EHEC O157:H7 TW14359	Sakai	EC4115	EDL933	O26:H11 Str. 11368	O103:H2 Str. 12009	O111:H- Str 11128	EPEC O55:H7 CB9615	O127:H6 Str E2348/69	ETEC H10407	E24377A	UMNK88	AIEC LF82	O83:H1 NR G857C	UM146	EAEC 042	55989	NMEC O7:K1 CE10	S88	IH E3034	UPEC UMN026	CLONE Di14	CLONE Di2	CFT073	IAI39	536	NA114	UTI89	AREC SMS-3-5	APEC01	ABU 83972
MC001	■	■	■	■	■	■	■	■	■																						
MC002	■	■	■	■	■	■	■	■																							
MC003	■	■	■	■				■								■		■			■				■				■		
MC004	■	■	■	■		■	■																								
MC005		■		■	■	■	■																								
MC006	■	■	■	■				■																							
MC007	■	■	■	■				■																							
MC008	■	■	■	■				■																							
MC009	■	■	■	■				■																							
MC010	■	■	■	■				■																							
MC011	■	■	■	■				■																							
MC012	■	■	■	■																											
MC013	■	■	■	■				■								■		■			■				■				■		
MC014	■	■	■	■																											
MC015	■	■	■	■				■																							
MC016	■	■	■	■				■																							
MC017	■	■	■	■	■		■						■																		
MC018	■	■	■			■		■		■																					
MC019	■	■	■	■																											
MC020	■	■	■	■				■		■		■				■					■								■		
MC021	■	■	■	■				■											■										■		
MC022	■	■	■	■				■								■		■			■				■				■		
MC023	■	■	■	■		■	■	■									■														
MC024	■	■	■	■				■																							

Figure 1S. The selected candidates are mainly present and conserved among EHEC strains. Gene variability and distribution analysis (by BLASTP) shows the presence/absence of the antigens in a panel of 31 *E. coli* InPEC and ExPEC complete genomes. Hits are represented by Black cells (sequence identity ≥ 80%, query coverage ≥ 90%) white cells represent gene absence or presence with a sequence identity <80% and query coverage < 90%. EHEC: enterohemorrhagic *E. coli*; EPEC: enteropathogenic *E. coli*; ETEC: enterotoxigenic *E. coli*; EIEC: enteroinvasive *E. coli*; EAEC: enteroaggregative *E. coli* (EAEC); NMEC: neonatal meningitis *E. coli;* UPEC: uropathogenic *E. coli*; AREC: antibiotic resistant *E. coli;* APEC avian pathogenic *E. coli*; ABU: Asymptomatic *E. coli*.

Chapter 2

As explained previously, the secretion of specific virulence factors is one of the main differences between non-pathogenic bacteria and those responsible for diseases. Such secreted virulence factors can be released in the extracellular milieu or on the cell surface and be involved in adhesion, biofilm formation and/or persistence of the bacteria in the host. Secretion systems encoded in the LEE locus and involved in pathogenicity are best characterized in EHEC and EPEC. However, regarding the secretome concept, other entities, such as the autotransporters, can participate in pathogenesis and colonization in animals, humans and food. As some autotransporters are exposed on the surface of cells, they can be good antigen candidates for immunological recognition and thus vaccine development. Immunoproteomic analysis has also been employed as a strategy for the identification of antigens able to generate an immune response in the host during the process of infection. During this analysis, autotransporters (AT) such as EtpA, Antigen 43 and TibA, were found to be reactive during the infection process. *In silico* tools have identified autotransporters in biofilm formation, such as pAT, antigen 43 and EatA, which are more exclusively found in ETEC or other pathotypes but are absent in commensal strains. Our own reverse vaccinology approach identified MC021 as a potent antigen. It has been further shown that recombinant Ag43 and pAT autotransporters increase fecal IgA and provide relative protection against ETEC intestinal colonization in immunized mice. Interestingly, these autotransporters are also recognized by sera from patients with ETEC diarrhea, confirming their expression during ETEC infections. For these reasons, the study and description of auotransporters has gained importance. Autotransporters correspond to T5aSS. They are modular proteins composed of (i) an N-terminal signal peptide, which is cleaved off after crossing of the inner membrane, (ii) a C-terminal translocation unit called translocator composed of a C-terminal β-barrel forming a pore through the outer membrane and an α-helix linker region, which (iii) links the β-barrel to the passenger domain. The passenger domain, corresponding to the functional domain of an autotransporter, can be cell surface-exposed or released in the extracellular milieu. Some passengers also exhibit an autochaperone (AC) domain, which is required for proper folding and secretion of the passenger through the outer membrane. However, the commonality of this domain among all autotransporters remains controversial. The aim of this second study carried out during this thesis was to explore the prevalence of ACs in autotransporters, using the unique model structure of an AC domain as revealed in IcsA, following structural alignment and phylogenetic analyses. By clarifying the prevalence of the AC in T5aSS, where it appeared to be solely associated with passengers with a β-helix structure, this study will provide information

supporting the idea that the AC structure could be used as a potential epitope target and/or for the development of new treatments.

Identification of the autochaperone domain in the Type Va secretion system (T5aSS): A prevalent feature of autotransporters with a β–helical passenger

Maricarmen ROJAS-LOPEZ [1,2,#], Mohamed A. ZORGANI [1,#], Lawrence A. KELLEY [3], Xavier BAILLY [4], Andrey V. KAJAVA [5], Ian R. HENDERSON [6], Fabio POLTICELLI [7,8], Mariagrazia PIZZA [2], Roberto ROSINI [2] and Mickaël DESVAUX [1,*].

[1] Université Clermont Auvergne, INRA, UMR454 MEDiS, 63000 Clermont-Ferrand.

[2] GlaxoSmithKline Vaccines, Via Fiorentina 1, 53100 Siena, Italy.

[3] Imperial College London, Structural Bioinformatics Group, London, United Kingdom.

[4] INRA, UR346 Epidémiologie animale, Saint Genès Champanelle, France.

[5] Centre de Recherches en Biologie cellulaire de Montpellier, UMR5237 CNRS, Institut de Biologie Computationnelle, Université Montpellier, 34293 Montpellier, France.

[6] University of Birmingham, Institute of Microbiology and Infection, Edgbaston, Birmingam, United Kingdom.

[7] Department of Sciences, Roma Tre University, Rome, Italy.

[8] National Institute of Nuclear Physics, Roma Tre Section, Rome, Italy.

[#] These authors contributed equally to this work and thus share first authorship.

*Corresponding author: Dr Hab Mickaël DESVAUX-LENÔTRE (Hab, PhD, Mag), INRA (Institut National de la Recherche Agronomique) Auvergne-Rhône-Alpes, Centre de Recherche de Clermont-Ferrand, UMR454 MEDiS (Microbiologie, Environnement Digestif, Santé), Site de Theix, F-63122 Saint-Genès Champanelle, France. Tel.: +33 (0)4 73 62 47 23, Fax: +33 (0)4 73 62 45 81, E-mail: mickael.desvaux@inra.fr.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 05 January 2018
doi: 10.3389/fmicb.2017.02607

Identification of the Autochaperone Domain in the Type Va Secretion System (T5aSS): Prevalent Feature of Autotransporters with a β-Helical Passenger

Maricarmen Rojas-Lopez[1,2†], Mohamed A. Zorgani[1†], Lawrence A. Kelley[3], Xavier Bailly[4], Andrey V. Kajava[5], Ian R. Henderson[6], Fabio Polticelli[7], Mariagrazia Pizza[2], Roberto Rosini[2] and Mickaël Desvaux[1]*

[1] Université Clermont Auvergne, INRA, UMR454 MEDiS, Clermont-Ferrand, France, [2] GSK, Siena, Italy, [3] Structural Bioinformatics Group, Imperial College London, London, United Kingdom, [4] Institut National de la Recherche Agronomique, UR346 Epidémiologie Animale, Saint Genès Champanelle, France, [5] CRBM UMR5237 CNRS, Institut de Biologie Computationnelle, Université Montpellier, Montpellier, France, [6] Institute of Microbiology and Infection, University of Birmingham, Birmingham, United Kingdom, [7] Department of Sciences, National Institute of Nuclear Physics, Roma Tre University, Rome, Italy

OPEN ACCESS

Edited by:
Daniela De Biase, Sapienza Università di Roma, Italy

Reviewed by:
Sophie Bleves, Aix-Marseille University, France
J. P. Van Ulsen, VU University Amsterdam, Netherlands

***Correspondence:**
Mickaël Desvaux
mickael.desvaux@inra.fr

† These authors have contributed equally to this work and shared first authorship.

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: 18 October 2017
Accepted: 14 December 2017
Published: 05 January 2018

Citation:
Rojas-Lopez M, Zorgani MA, Kelley LA, Bailly X, Kajava AV, Henderson IR, Polticelli F, Pizza M, Rosini R and Desvaux M (2018) Identification of the Autochaperone Domain in the Type Va Secretion System (T5aSS): Prevalent Feature of Autotransporters with a β-Helical Passenger. Front. Microbiol. 8:2607. doi: 10.3389/fmicb.2017.02607

Autotransporters (ATs) belong to a family of modular proteins secreted by the Type V, subtype a, secretion system (T5aSS) and considered as an important source of virulence factors in lipopolysaccharidic diderm bacteria (archetypical Gram-negative bacteria). While exported by the Sec pathway, the ATs are further secreted across the outer membrane via their own C-terminal translocator forming a β-barrel, through which the rest of the protein, namely the passenger, can pass. In several ATs, an autochaperone domain (AC) present at the C-terminal region of the passenger and upstream of the translocator was demonstrated as strictly required for proper secretion and folding. However, considering it was functionally characterised and identified only in a handful of ATs, wariness recently fells on the commonality and conservation of this structural element in the T5aSS. To circumvent the issue of sequence divergence and taking advantage of the resolved three-dimensional structure of some ACs, identification of this domain was performed following structural alignment among all AT passengers experimentally resolved by crystallography before searching in a dataset of 1523 ATs. While demonstrating that the AC is indeed a conserved structure found in numerous ATs, phylogenetic analysis further revealed a distribution into deeply rooted branches, from which emerge 20 main clusters. Sequence analysis revealed that an AC could be identified in the large majority of SAATs (self-associating ATs) but not in any LEATs (lipase/esterase ATs) nor in some PATs (protease autotransporters) and PHATs (phosphatase/hydrolase ATs). Structural analysis indicated that an AC was present in passengers exhibiting single-stranded right-handed parallel β-helix, whatever the type of

β-solenoid, but not with α-helical globular fold. From this investigation, the AC of type 1 appears as a prevalent and conserved structural element exclusively associated to β-helical AT passenger and should promote further studies about the protein secretion and folding via the T5aSS, especially toward α-helical AT passengers.

Keywords: protein secretion system, Autotransporters, Type V secretion system, Outer membrane proteins, Protein translocation, Autochaperone domain, diderm-LPS Gram-negative bacteria

INTRODUCTION

Bacteria can secrete proteins by numerous molecular machineries. In this field, it is of key importance to differentiate export (protein transport across the cytoplasmic membrane) from secretion (protein transport from inside to outside the cell) systems (Desvaux et al., 2004b; Economou et al., 2006), especially in lipopolysaccharidic diderm (LPS-diderm) bacteria (archetypical Gram-negative bacteria) (Desvaux et al., 2009; Sutcliffe, 2010; Chagnot et al., 2013). While Sec (Secretion) and Tat (Twin-arginine translocation) systems constitute the two major export pathways, nine protein secretion systems numbered from Type I to Type IX (T1SS-T9SS) are currently recognised in the LPS-diderm bacteria, which enable protein transport across the outer most biological membrane. Among them, the T5SS most certainly secrete the most diverse range of effectors, thus constituting a premium source of virulence factors (Henderson and Desvaux, 2004; Henderson et al., 2004). The T5SS is currently subdivided into the (i) autotransporters (T5aSS), (ii) two-partner passenger-translocators (T5bSS), (iii) trimeric autotransporters (T5cSS), (iv) hybrid autotransporters (T5dSS), and (v) inverted autotransporters (T5eSS) (Desvaux et al., 2003; Leo et al., 2012).

The T5SS can be broadly defined as protein transport across the asymmetric LPS-containing outer membrane (OM) *via* a β-barrel to complete the secretion, which is first initiated by protein export *via* the Sec machinery for cytoplasmic inner membrane (IM) transit (Henderson et al., 2004; Leo et al., 2012). Nonetheless, several periplasmic chaperones, the BAM (β-barrel assembly machinery) and TAM (translocation and assembly module) complexes take part to the OM secretion process (Leyton et al., 2012; Selkrig et al., 2012). Regarding the T5aSS, the autotransporters (ATs) have a modular architecture constituted of three major regions, (i) a N-terminal signal peptide (SP), (ii) a central passenger, and (iii) a C-terminal translocator (Desvaux et al., 2004a; Drobnak et al., 2015). The SP targets the proteins to the IM before being cleaved off after export *via* Sec. Some autotransporter SPs exhibit a highly conserved domain called ESPR (Extended Signal Peptide Region) (Desvaux et al., 2006), which influence IM and OM translocation but whose exact function remains unclear (Desvaux et al., 2007; Jong and Luirink, 2008). The passenger is secreted across the OM and corresponds to the effector, which is either displayed at the bacterial cell surface or further released into the extracellular milieu. The passengers are generally believed to exhibit a β-helical structure (Jenkins et al., 1998; Kajava and Steven, 2006) but this is not systematic, e.g. EstA (Esterase Autotransporter) has a globular fold dominated by α-helices and loops, which is regarded as a general feature in the lipase/esterase ATs (LEATs) (Van Den Berg, 2010; Celik et al., 2012). In the β-helical passenger of ATs, different types of a β-solenoid motif are currently recognised as either displaying a triangular or L-shape coil cross-sections (Kajava and Steven, 2006). The translocator forms the translocation unit (TU) composed of an α-helical linker and a β-barrel domain, through which the passenger is transported across the OM (Oomen et al., 2004).

The investigation of secretion and folding of BrkA (*Bordetella* resistance to killing) evidenced the importance of an autochaperone (AC) domain localised at the C-terminal region of the passenger was evidenced (Oliver et al., 2003). Besides BrkA, requirement of the AC for proper passenger secretion and cell-surface folding was supported by several investigations in AIDA-I (*Escherichia coli* Adhesin Involved in Diffuse Adherence I) (Berthiaume et al., 2007), EspP (Extracellular serine protease precursor) (Velarde and Nataro, 2004), Hbp (Hemoglobin-binding protease) (Soprova et al., 2010), IcsA (Intra-cellular spread protein A) (May and Morona, 2008), Pet (Plamid-encoded toxin) (Dutta et al., 2003) and Ssp (*Serratia marcescens* serine protease) (Ohnishi et al., 1994). When mutated, secretion of BrkA could be rescued with the AC supplied in *trans* (Oliver et al., 2003), which was also demonstrated in Ssp (Ohnishi et al., 1994), EspP and Pet (Dutta et al., 2003). This suggested that the AC provides a template-induced folding mechanism for the passenger. More recently, the crystal structure of the AC of IcsA was resolved and this clearly appeared to exhibit a characteristic β-sandwich fold (Kuhnel and Diezmann, 2011). However, considering it was functionally characterised and identified only in a handful of autotransporters, suspicion recently fell on the commonality and conserved nature of the AC as a sequence element (Drobnak et al., 2015). Still, the identification of this domain could be limited by classical BLAST search due to sequence divergence (Altschul and Koonin, 1998).

This prompted us to investigate the prevalence of the AC in the T5aSS. Taking advantage of the resolved three-dimensional structure of the AC from IcsA, this domain was searched by structural alignment, first, among all passengers experimentally resolved by crystallography, before searching against a database of well-defined and genuine ATs. From there, phylogenetic analysis revealed the organisation of the AC family, whereas proteogenomic analysis pinpointed that the AC was systemically associated with passengers exhibiting a β-helix fold but not a globular fold like EstA. Taken together, the AC appears as a conserved domain exclusively present in ATs with a β-helical passenger.

MATERIALS AND METHODS

Structural Alignment

Superimposition of one structure against another was performed using PyMOL v1.7.4. The atomic coordinates and structure information of the proteins of interest were recovered from the Protein Data Bank (PDB) (Berman et al., 2000a,b). The structures were aligned to minimise the RMSD (Root Mean Square Deviation) between the aligned atoms (C-alfas). For multiple sequence alignment, a special mode of T-Coffee (Notredame et al., 2000; Di Tommaso et al., 2011; Magis et al., 2014) was used to incorporate structural information, i.e. Expresso (Armougom et al., 2006), which is an extension of 3D-Coffee where structure based alignment is used as a template for realigning the original sequences, which results in a structure-based multiple sequence alignment (O'sullivan et al., 2004). The alignment was then visualised with ENDscript, which combines both primary sequence and secondary structure alignment (Gouet et al., 2003; Robert and Gouet, 2014).

Search for Domain Homologs

The AC domains were also identified in autotransporters by submitting passenger sequences to Phyre v2.0 for automated modeling. The dataset included the 1523 well-defined and genuine autotransporters identified by the twin-HMM autotransporter procedure designed by Celik et al. (2012). In parallel, using the refined AC structures here defined from structural alignments of IcsA, EspP, Hbp, Pet, pertactin P69, Hap, and IgA1 (PDB files provided as Supplementary Materials), the presence of the AC in other autotransporters was searched using Phyre in reverse, i.e., BackPhyre (Kelley and Sternberg, 2009; Kelley et al., 2015). To achieve a high degree of reliability with respect to the predicted domain fold and modeling of the protein core at high accuracy (<4 Å RMSD from native, true structure), only structural matches with a high level of confidence ($>90\%$) were considered.

Phylogenetic Analysis of Protein Sequences

The ACs identified by structural alignment with Phyre/BackPhyre were aligned with T-Coffee in the expresso mode, using PDB files restricted to the AC domains of IcsA (PDB: 3ML3; D_{606}–L_{720}), EspP (PDB: 3SZE; D_{869}–I_{979}), Hbp (PDB: 1WXR; N_{948}–L_{1056}), Pet (PDB: 4OM9; N_{865}–I_{974}), P69 (PDB: 1DAB; D_{444}–L_{556}), Hap (PDB: 3SYJ; D_{830}–L_{964}), and IgA1 (PDB: 3H09; D_{865}–L_{977}) to seed the alignment. A BioNJ tree (Gascuel, 1997) based on observed divergences between pairs of sequences was obtained using SplitsTree (Kloepper and Huson, 2008). The most relevant clusters, i.e., monophyletic groups or clades, were identified and selected based on splits showing bootstrap values above 80% over 1,000 pseudo-replicates.

Identification of Functional Motifs and Secondary Structures

Functional motifs were searched using InterProScan v5.22 (Jones et al., 2014) and interrogating InterPro (IPR) v61.0 database (Finn et al., 2017), which includes CATH-Gene3D v4.1 (Sillitoe et al., 2015; Lam et al., 2016), CDD (Marchler-Bauer et al., 2017), MobiDB v2.3.2014.07 (Potenza et al., 2015), HAMAP (Pedruzzi et al., 2015), PANTHER v11.1 (Mi et al., 2017), Pfam v30.0 (Finn et al., 2016), PIRSF (Wu et al., 2004), PRINTS (Attwood et al., 2003), ProDom v2012.1 (Servant et al., 2002), PROSITE v2017.01 (Sigrist et al., 2013), SFLD (Akiva et al., 2014), SMART v7.0 (Schultz et al., 1998), SUPERFAMILY v1.75 (Wilson et al., 2009), TIGRFAMs v15.0 (Haft et al., 2003). Besides structure modeling using Phyre v2, β-helix folds were predicted with BetaWrap, using rung profile (3–7 rungs) (Bradley et al., 2001), as well as PfScan to identify the types of β-solenoid motif (Kajava and Steven, 2006).

RESULTS

Identification of the Autochaperone Domain in Autotransporter Passengers with a Resolved Three-Dimensional Structure

While the autochaperone (AC) was experimentally identified in BrkA, AIDA-I, EspP, Hbp, Pet, and Ssp, neither the structure of the passenger of BrkA, AIDA-I nor Ssp has been experimentally resolved by crystallography, or by any other mean, contrary to EspP (PDB: 3SZE) (Khan et al., 2011), Hbp (PDB: 1WXR) (Otto et al., 2005), and Pet (PDB: 40M9) (Domingo Meza-Aguilar et al., 2014) passengers. Using the structure of the AC of IcsA as reference (AC^{IcsA}; PDB: 3ML3) (Kuhnel and Diezmann, 2011), identification of the AC in the C-terminal part of EspP, Hbp, and Pet was first attempted. While the resolution of the secondary structure of Pet was low in the C-terminal region of the passenger (PDB: 40M9) and the RMSD was slightly high for Hbp when superimposed to AC^{IcsA} (RMSD $>$ 4 Å over 54 atoms), the 3D-structure of the C-terminal part of EspP (D_{869}–I_{979}) superimposed onto residues D_{606}–L_{720} of IcsA (**Figure 1**) with RMSD of 2.51 Å over 43 atoms (**Table 1**). The structural region here identified as the AC^{EspP} completely agrees with previous experimental investigations by functional mutagenesis where the AC was identified within the 821–997 region of EspP (Dutta et al., 2003; Velarde and Nataro, 2004; Skillman et al., 2005; Brockmeyer et al., 2009). Similarly, AC^{Pet} (N_{865}–I_{974}) was here identified within the 819–992 region of Pet, functionally characterised as an autochaperone (Dutta et al., 2003) and within the 950–1,048 region for the AC^{Hbp} (N_{948}–L_{1056}) (Soprova et al., 2010). Using the AC^{IcsA} structure (D_{606}–L_{720}), this domain was structurally aligned within the C-terminal region of all other crystallised autotransporter passengers, namely pertactin P69 (PDB: 1DAB), Hap (PDB: 3SYJ), IgA1 (PDB: 3H09), EstA (PDB: 3KVN), Ag43 (PDB: 4KH3), and VacA (PDB: 2QV3). Except for EstA, Ag43, and VacA, an AC could be identified in all other resolved AT passengers. Following structure alignment, the RMSD varied between 1.51 Å for AC^{IcsA} vs. AC^{P69} and 3.40 Å for AC^{IcsA} vs. AC^{IgA1} (**Table 1**). Once identified in each passenger, these AC domains were further structurally aligned one with another and appeared to superimpose with RMSD ranging from 0.44 to 2.97 Å for AC^{Hbp} vs. AC^{EspP} and for AC^{P69} vs. AC^{Pet} respectively (**Table 1**). Of note, the AC^{Hbp} superimposed to the

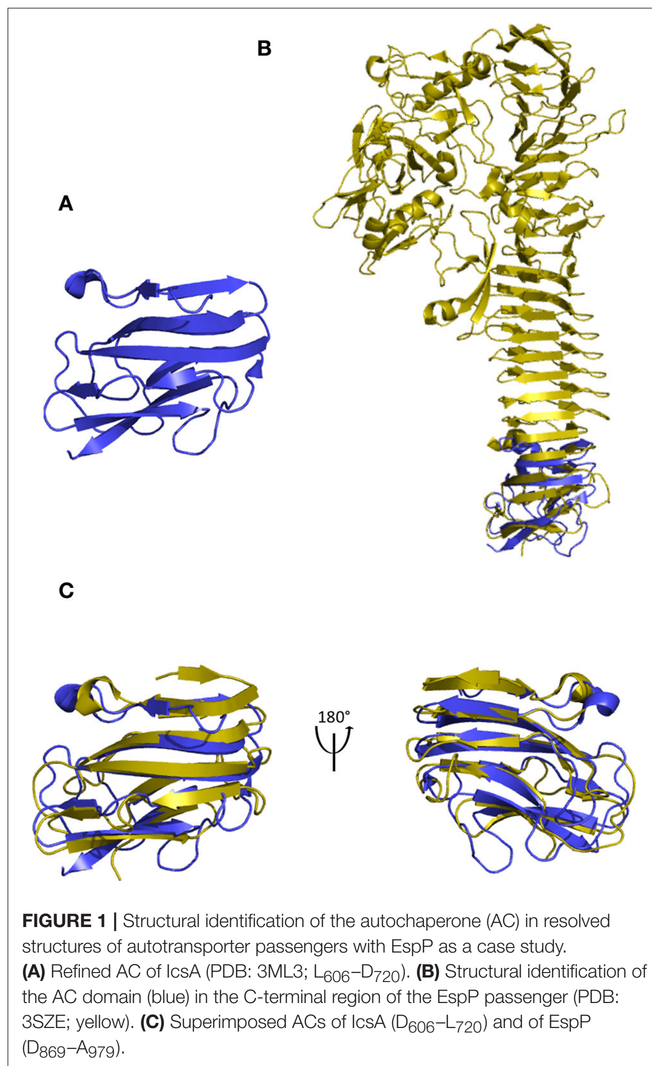


FIGURE 1 | Structural identification of the autochaperone (AC) in resolved structures of autotransporter passengers with EspP as a case study. **(A)** Refined AC of IcsA (PDB: 3ML3; L_{606}–D_{720}). **(B)** Structural identification of the AC domain (blue) in the C-terminal region of the EspP passenger (PDB: 3SZE; yellow). **(C)** Superimposed ACs of IcsA (D_{606}–L_{720}) and of EspP (D_{869}–A_{979}).

TABLE 1 | RMSD values of superimposed AC domains present in the experimentally resolved autotransporters three-dimensional structures.

RMSD[a]	AC^{EspP}	AC^{Hbp}	AC^{P69}	AC^{Hap}	AC^{IgA1}	AC^{Pet}
AC^{IcsA}	2.51 (43)	6.10 (67)	2.69 (59)	1.88 (81)	3.39 (74)	3.97 (97)
AC^{Pet}	1.14 (83)	1.42 (70)	2.97 (33)	1.71 (54)	2.05 (54)	–
AC^{IgA1}	2.48 (72)	1.11 (48)	2.02 (74)	0.54 (66)	–	–
AC^{Hap}	2.18 (62)	1.27 (55)	0.89 (69)	–	–	–
AC^{P69}	2.84 (69)	1.31 (54)	–	–	–	–
AC^{Hbp}	0.44 (94)	–	–	–	–	–

[a] *RMSD (Root Mean Square Deviation) values (Å) with the number of superimposed atoms (in brackets) for each AC domains superimposed one with another and identified from the known three-dimensional structures of autotransporters IcsA (PDB: 3ML3; AC: D_{606}–L_{720}), EspP (PDB: 3SZE; AC: D_{869}–I_{979}), Hbp (PDB: 1WXR; AC: N_{948}–L_{1056}), Pet (PDB: 4OM9; AC: N_{865}-I_{974}), pertactin P69 (PDB: 1DAB; AC: D_{444}–L_{556}), Hap (PDB: 3SYJ; AC: D_{830}–L_{964}), and IgA1 (PDB: 3H09; AC: D_{865}–L_{977}).*

AC of these other ATs with RMSD systematically lower than 2 Å (**Table 1**). With a size ranging from 109 to 135 amino acid (a.a.) residues, the AC domains displayed a conserved structure forming coils of parallel and anti-parallel β-sheets and a couple of short α-helices (**Figure 2A**). For IgA1 and Pet, however, the resolution of the structure in this region was not high enough to provide information on the secondary structures. Sequence similarities of the identified AC were further confirmed following multiple sequence alignment incorporating structural information, where regions with β-strands or α-helices aligned one another in the different ACs (**Figure 2B**).

Distribution of the Autochaperone among the T5aSS

To determine whether similar AC structures are present in other autotransporters, the well-defined ACs here identified from solved tertiary structures of passengers, namely AC^{IcsA} (D_{606}–L_{720}), AC^{EspP} (D_{869}–I_{979}), AC^{Hbp} (N_{948}–L_{1056}), AC^{P69} (D_{444}–L_{556}), and AC^{Hap} (D_{830}–L_{964}), were used as queries to search for similar structures using BackPhyre. In parallel, the AC domains were also identified by submitting the sequences corresponding to the passenger domain to Phyre for automated modeling. For data mining, the dataset was constituted of the well-defined and genuine ATs identified by Celik et al. (2012). Out of these 1523 sequences of autotransporters, an AC could be identified in 708 of them with confidence levels higher than 90.0%, which corresponded to 429 non-redundant and distinct sequences (Supplementary Material Table 1S). These AC domains have an average size of about 110 a.a. residues (mediane = 112 and mode = 115 a.a.) and were essentially present toward the C-terminal region of the passengers, adjacent to the translocator region at an average distance of about 100 a.a. residues upstream (mediane = 84 and mode = 67 a.a.). With respect to the ATs for which an AC was experimental identified but no structural information was available, an AC could indeed be identified in the C-terminal region of the BrkA and AIDA-I passengers but not Ssp. Regarding ATs for which the passenger structure was resolved but an AC domain could not be identified in the first instance, an AC was finally identified in the C-terminal part of the Ag43 passenger (position 525-638), i.e., a region that has not been crystallised (Heras et al., 2014), but this was not the case either for EstA or for VacA. Following multiple sequence alignment incorporating structural information, phylogenetic analysis revealed that the AC domains distribute into deeply rooted branches, from which 20 main clusters emerged (named according to cyrillic alphabet with phonetics in brackets), namely А [a], Б [b], В [v], Г [g], Д [d], Е [je], Ж [ʒ], З [z], И [i], Й [j], К [k], Л [l], М [m], Н [n], О [o], П [p], Р [r], С [s], Т [t], and У [u] (**Figure 3**), where clusters А and Б form the largest groups hosting some autotransporter members of Iba (Inducible *Bartonella* autotransporter) (Eicher and Dehio, 2012) and AutA (Auto-transporter A) groups respectively (Ait-Tahar et al., 2000). Clusters Г, Д, З, И, Й, К, Л, О, С, and У harbor ACs from autotransporters that have not been characterised yet. The resolved AC domains from IcsA, EspP, Pet, or pertactin P69 were not found in clusters but in deeply rooted branches, whereas the ACs from Hbp, IgA1, and Hap were part of clusters М, Р, and Т respectively (**Figure 3**). Considering other characterised autotransporters, clusters В, Е, Ж, Н, and П harbor ACs from EhaC, Ag43, AIDA-I, EhaD, and YcgV autotransporter members

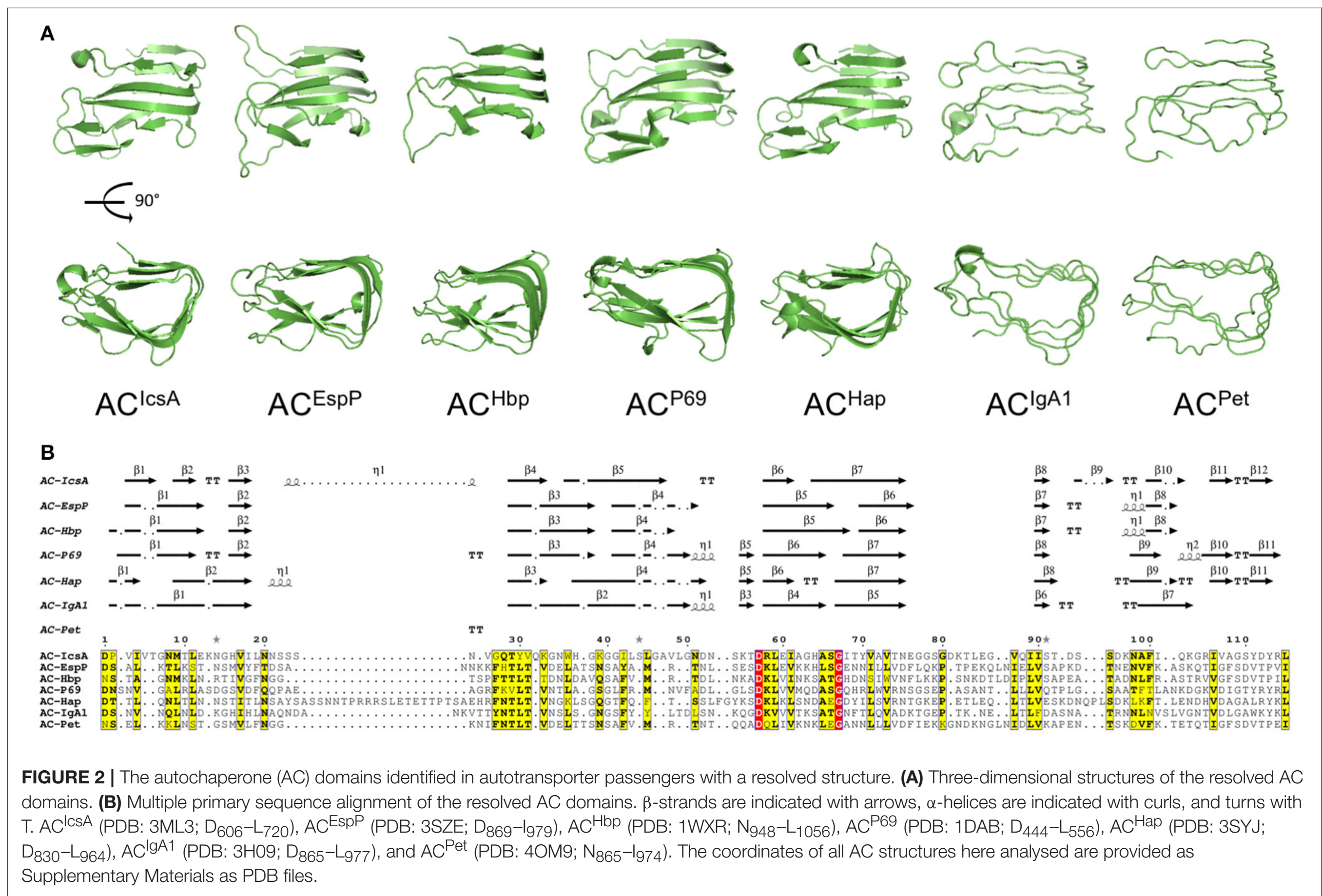


FIGURE 2 | The autochaperone (AC) domains identified in autotransporter passengers with a resolved structure. **(A)** Three-dimensional structures of the resolved AC domains. **(B)** Multiple primary sequence alignment of the resolved AC domains. β-strands are indicated with arrows, α-helices are indicated with curls, and turns with T. AC^{IcsA} (PDB: 3ML3; D_{606}–L_{720}), AC^{EspP} (PDB: 3SZE; D_{869}–I_{979}), AC^{Hbp} (PDB: 1WXR; N_{948}–L_{1056}), AC^{P69} (PDB: 1DAB; D_{444}–L_{556}), AC^{Hap} (PDB: 3SYJ; D_{830}–L_{964}), AC^{IgA1} (PDB: 3H09; D_{865}–L_{977}), and AC^{Pet} (PDB: 4OM9; N_{865}–I_{974}). The coordinates of all AC structures here analysed are provided as Supplementary Materials as PDB files.

(Vo et al., 2017), whereas the ACs from BrkA and Pet were found in deeply rooted branches.

The AC Is Exclusively Associated with Autotransporters Exhibiting β-Helical Passengers

Looking for a correlation with the presence of the AC, functional genomic analysis of the ATs was performed (Supplementary Material Table 1S). First, an ESPR (IPR024973) could be predicted in the SP region of 134 ATs; in 84% of these an AC was also identified. Based on the functional motifs identified in the passengers, the ATs could be further classified into 6 main and distinct functional categories, *i.e.* the (i) protease autotransporters (PATs), (ii) self-associating autotransporters (SAATs), (iii) phosphatase/hydrolase autotransporters (PHATs), (iv) lipase/esterase autotransporters (LEATs), (v) vacuolating autotransporters (VATs), and (vi) autotransporters of unknown function. The PATs, like Pet, Ssp, or EspP, could belong to different peptidase families, essentially the (i) serine peptidases S1 (IPR001314), S6 (IPR000710) or S8/S53 (IPR000209), (ii) cysteine peptidase C1 (IPR025660), or (iv) metallopeptidases M10 (IPR011049) or M28 (IPR007484). SAATs systematically feature an adhesion domain of ATs (IPR030930) like AIDA or Ag43. PHATs mainly belong to either the (i) phosphatidic phosphatase (IPR000326), (ii) tyrosine phosphatase (IPR029021), or (iii) glycoside hydrolase (IPR002772) family. LEATs, like EstA, possess GSDL lipase/esterase (IPR001087) and/or SGNH esterase (IPR013830) domains. Like VacA, the VATs systemically exhibit a vacuolating cytotoxin domain (IPR004311). The most striking observation was that no AC could be identified in any of the LEATs or VATs, whereas it was identified in the large majority of the SAATs (83%). Among the PATs and PHATs, an AC was present or absent from some functional subcategories, e.g. the M28 metalloproteases or the tyrosine phosphatases. Regarding the ATs for which no function could be inferred, homology to pectin lyase fold (IPR011050), pertactin P69 (IPR004899; IPR003991) and/or P22 tailspike protein (IPR012332) was predicted in the passengers. These regions correspond to β-helix topologies as encountered in most ATs (Jenkins et al., 1998), including the SAATs. Interestingly, all passengers belonging to the LEATs, as well as to the M28 metalloprotease or the tyrosine phosphatase ATs, exhibit α-helical folds for which no AC could be identified. Of note, the passenger of Ssp was predicted to essentially display only α-helices, like EstA, for which no AC structure could be identified either (**Figure 4**). Following structural modeling of the passengers, it appeared that the AC is systematically associated with passenger exhibiting β-helix folds, e.g. as observed in the passengers with a resolved structure, namely Ag43, EspP,

FIGURE 3 | Phylogenetic tree of the AC domains in the T5aSS. Sequences of the AC domain were identified among the non-redundant well-defined dataset of 1523 autotransporters (Celik et al., 2012) following structural search using Phyre/BackPhyre. For legibility, the redundant and very closely related AC sequences were removed from the tree but are readily available in Table 1S. The scale bar represents the evolutionary distance, i.e., the number of substitutions per site.

Hap, Hbp, IcsA, IgA1, pertactin P69, and Pet, or in predicted β-helical passengers where an AC was experimentally identified, namely BrkA and AIDA-I. Despite predominantly exhibiting a single-stranded right-handed parallel β-helix in the passenger, no AC could be identified in VacA or any VATs, indicating that such as a fold is not always associated with an AC domain. Considering the oval (O), triangular (T) or L-shaped coil cross-sections, 16 different β-solenoids are currently recognised (Kajava and Steven, 2006). As expected, no O1 or O2 repeat could be identified in any of the AT passengers as these coils occur upon oligomerisation, as observed in trimeric ATs belonging to the T5cSS. Besides L1, L3, T4, T5, and T6-type

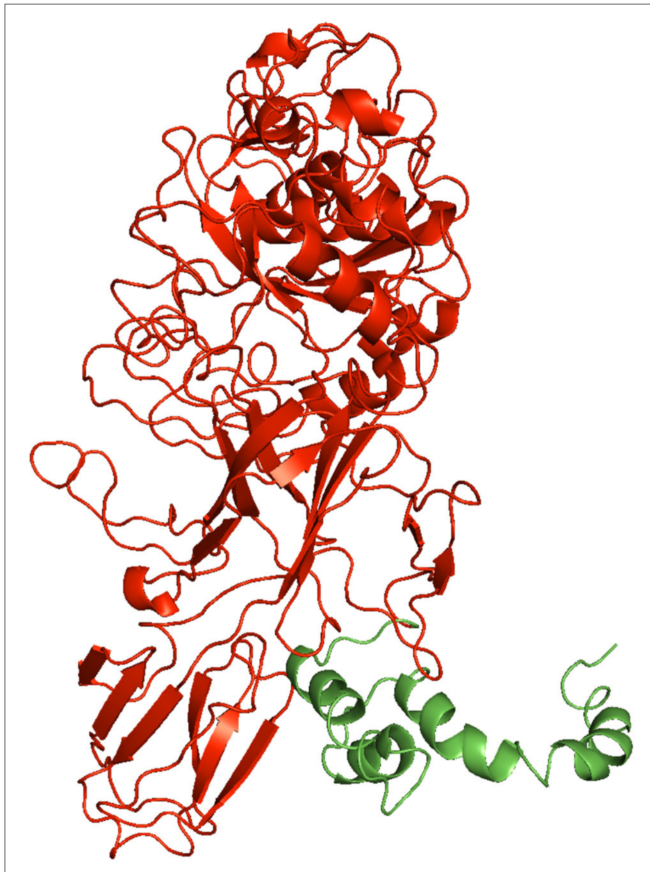


FIGURE 4 | Three-dimensional structure modeling of the passenger of Ssp. The passenger sequence of Ssp (A_{28}–G_{716}) was submitted as query to Phyre using intensive mode. The structure model covers the 37–601 region (72% coverage of the query sequence) with a confidence level of 100.0% with the single highest scoring template (Subtilase family, PDB: 1R6V). The green and red colors depict the regions corresponding to the AC domain and functional domain of the passenger respectively. The structure of the AC domain was essentially modeled *ab initio* and must be regarded as unreliable. The coordinates of the Ssp passenger model are provided as Supplementary Material as a PDB file.

β-solenoids previously reported as specific to AT passengers, some coils originally considered as specific to the T5bSS, i.e., the TPS (two-partner system), were also identified and encountered in association with an AC, namely the L2, T1, and T8 coils. Rather than being associated to some specific functions, the AC thus appears exclusively associated with β-helical passengers.

DISCUSSION

While the function of the AC for proper passenger secretion and folding on the cell-surface has been well demonstrated in several ATs, its commonality in the T5aSS was recently questioned. This ambiguity partly results from the difficulty in identifying the AC in uncharacterised ATs following similarity search by sequence alignment. Following a structural approach, the AC domain was here identified, first in other ATs whose passenger structure was resolved and subsequently in ATs from a recognised reference dataset (Celik et al., 2012). It must be stressed that the prediction was based on similarity of the ACs, including functionally characterised ACs, namely EspP (Velarde and Nataro, 2004), Hbp (Soprova et al., 2010), IcsA (May and Morona, 2008), and Pet (Dutta et al., 2003). It further provides the proof-of-principle that a predicted AC can indeed be functional since the AC structure was here identified in BrkA and AIDA-I, which AC domains were functionally demonstrated but not structurally resolved (Oliver et al., 2003; Berthiaume et al., 2007). It is worth mentioning that some other regions in the C-terminal half of the β-helical passengers could contribute to folding and secretion (Drobnak et al., 2015); They might function in conjunction with the AC of ATs that have them. Nevertheless and contrary to what previously believed (Drobnak et al., 2015), this investigation clearly demonstrates that the AC corresponds to a conserved structural element present in the passenger of numerous, but not all, ATs. Besides, the AC appeared to be systematically and exclusively associated to passengers exhibiting single-stranded right-handed parallel β-helix, whenever the coils belong to the L1, L2, L3, T1, T4, T5, T6, or T8-type β-solenoids. The C-terminal region of the passenger, encompassing the AC, is involved in the initiation of folding of the passenger and prevent its unfolding once formed (Junker et al., 2006; Soprova et al., 2010; Renn et al., 2012; Baclayon et al., 2016). In Hbp, the stacking of aromatic residues was found to be important for its folding and stability (Baclayon et al., 2016). Considering the current proposed model mechanism, where the AC provides the first β-helical rung to promote folding of the passenger at the cell surface after emerging from the translocator, it makes sense that the AC is prevalent and even restricted to passenger with a β-helical architecture. A priori, it is almost impossible to predict the chirality of a β-helical structure since both right- and left-handed β -helices can be expected (Kajava et al., 2001). Because it triggers the right-handed arrangement, the presence of an AC domain unambiguously indicates that the upstream region has a right-handed β-helix. While the AC is generally located at the vicinity of the translocator, it is not always the case as already reported in IgA1, where it is present at another C-terminus part of the passenger that results from post-secretional processing (Oliver et al., 2003). This observation applies to all ACs identified in cluster P, which regroup members of the IgA1 AT family. While we found additional ACs at a significant distance from the translocator (Table 1S), more thorough investigations are needed to demonstrate whether they are subjected to similar processing.

While all ACs were found exclusively associated to β-helical passengers, not all passengers with a β-helix fold seem to possess an AC. Unexpectedly, an AC domain was not identified in any of the VATs, which are still predicted to have a single-stranded right-handed parallel β-helix just like in VacA (Gangwer et al., 2007). This suggests that either an AC with a different fold is present in the VATs and could not be identified by our approach or the secretion and folding of VacA-like ATs do not require any kind of AC. Thus, although the co-existence of β-helical domain and AC has generally been assumed in ATs, this work provides for the first time evidence that this relationship is not always straightforward.

Interestingly, no AC could be identified in the AT passengers predicted to have an EstA like globular fold, dominated by α-helices and loops (Van Den Berg, 2010). Unexpectedly, the Ssp passenger was predicted with a high confidence level (100.0%) to display some α-helical folds where no AC structure could be identified (**Figure 4**). Actually, it is in this AT that a region with an intramolecular chaperone function was for the first time reported (Ohnishi et al., 1994). Interestingly, the region corresponding to the AC domain displayed an α-fold and not a β-fold like for the AC presently identified. While the secondary structures in α-helices are predicted with a high confidence level in the S_{646}–G_{716} region reported as a functional AC (Ohnishi and Horinouchi, 1996), it must be stressed the tertiary structure of this region is essentially modeled *ab initio* and is unreliable; Intensive structural modeling was undertaken but failed to provide any significant tertiary structure prediction for this AC domain in Ssp. With no crystallographic data available, the structural nature of this region remains unknown but clearly differs from that of the ACs here reported. Besides stressing the need to experimentally resolve the three-dimensional structure of the AC^{Ssp}, this result pinpoints the need to experimentally determine the fold of passengers with architectures alternative to the single-stranded right-handed parallel β-helix. So far, EstA is the only AT with an α-helical passenger, which structure has been experimentally resolved (Van Den Berg, 2010). Besides the LEATs, however, such α-helical fold would occur in some PHATs and PATs (e.g. Ssp) as well. With such a gap of knowledge, we can only hypothesise on the possibility to have an alternative AC of type 2 (AC-2) with an α-fold associated to α-helical passenger as suggested by the investigation in Ssp; this would make the pair with the AC of type 1 (AC-1) presently investigated, which has a β-fold and is associated to β-helical passenger. Beyond the mechanistic aspects of the T5aSS, such information about the importance of these AC-1 and AC-2 in secretion and folding of the passenger is of great importance for biotechnological applications, e.g. aiming at efficiently expressing heterologous proteins for cell-surface display via the T5aSS (Van Ulsen et al., 2014; Nicolay et al., 2015), or for biomedical applications, e.g. aiming at blocking the AT secretion or folding to mitigate the virulence levels of pathogenic bacteria (Bondarenko et al., 2002; Wells et al., 2007).

AUTHOR CONTRIBUTIONS

MR-L, MZ, and MD conceived and designed the experiments. MR-L, MZ, LK, XB, and AK performed the experiments and data acquisition. MR-L, MZ, LK, XB, AK, IH, FP, MP, RR, and MD analysed and interpreted the data. MR-L, LK, XB, AK, FP, MP, RR, and MD contributed to materials and analysis tools. MR-L, MZ, LK, XB, AK, IH, FP, MP, RR, and MD wrote the article, including drafting and revising critically the manuscript for important intellectual content.

ACKNOWLEDGMENTS

This work was supported in part by INRA (Institut National de la Recherche Agronomique) and by EU FP7 Marie Curie Actions ITN (Initiale Training Network) EID (European Industrial Doctorate) DIS*Co* (A multidisciplinary Doctoral Industrial School on novel preventive strategies against *E. Coli* infections; http://www.discoproject.eu/; n°FP7-PEOPLE-607611) project. MR-L is a Marie Curie Ph.D. Research Fellow granted by ITN EID DIS*Co* performed at GSK and INRA. The authors are grateful to the INRA MIGALE bioinformatics platform (http://migale.jouy.inra.fr) for providing computational resources.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02607/full#supplementary-material

REFERENCES

Ait-Tahar, K., Wooldridge, K. G., Turner, D. P., Atta, M., Todd, I., and Ala'aldeen, D. A. (2000). Auto-transporter a protein of *Neisseria meningitidis*: a potent CD4+ T-cell and B-cell stimulating antigen detected by expression cloning. *Mol. Microbiol.* 37, 1094–1105. doi: 10.1046/j.1365-2958.2000.02061.x

Akiva, E., Brown, S., Almonacid, D. E., Barber, A. E. II., Custer, A. F., Hicks, M. A., et al. (2014). The structure-function linkage database. *Nucleic Acids Res.* 42, D521–D530. doi: 10.1093/nar/gkt1130.

Altschul, S. F., and Koonin, E. V. (1998). Iterated profile searches with PSI-BLAST–a tool for discovery in protein databases. *Trends Biochem. Sci.* 23, 444–447. doi: 10.1016/S0968-0004(98)01298-5

Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., et al. (2006). Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res.* 34, W604–W608. doi: 10.1093/nar/gkl092

Attwood, T. K., Bradley, P., Flower, D. R., Gaulton, A., Maudling, N., Mitchell, A. L., et al. (2003). PRINTS and its automatic supplement, prePRINTS. *Nucleic Acids Res.* 31, 400–402. doi: 10.1093/nar/gkg030

Baclayon, M., Ulsen, P., Mouhib, H., Shabestari, M. H., Verzijden, T., Abeln, S., et al. (2016). Mechanical unfolding of an autotransporter passenger protein reveals the secretion starting point and processive transport intermediates. *ACS Nano* 10, 5710–5719. doi: 10.1021/acsnano.5b07072

Berman, H. M., Bhat, T. N., Bourne, P. E., Feng, Z., Gilliland, G., Weissig, H., et al. (2000a). The Protein Data Bank and the challenge of structural genomics. *Nat. Struct. Biol.* 7(Suppl, 1), 957–959. doi: 10.1038/80734

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000b). The protein data bank. *Nucleic Acids Res.* 28, 235–242. doi: 10.1093/nar/28.1.235

Berthiaume, F., Rutherford, N., and Mourez, M. (2007). Mutations affecting the biogenesis of the AIDA-I autotransporter. *Res. Microbiol.* 158, 348–354. doi: 10.1016/j.resmic.2007.02.006

Bondarenko, V. M., Mavziutov, A. R., and Agapova, O. V. (2002). Serine proteases of Gram-negative bacteria: structure, mechanisms of secretion, biological activity. *Zh. Mikrobiol. Epidemiol. Immunobiol.* 6, 80–85.

Bradley, P., Cowen, L., Menke, M., King, J., and Berger, B. (2001). BETAWRAP: successful prediction of parallel beta -helices from primary sequence reveals an association with many microbial pathogens. *Proc. Natl. Acad. Sci. U.S.A.* 98, 14819–14824. doi: 10.1073/pnas.251267298

Brockmeyer, J., Spelten, S., Kuczius, T., Bielaszewska, M., and Karch, H. (2009). Structure and function relationship of the autotransport and proteolytic activity

of EspP from Shiga toxin-producing *Escherichia coli*. *PLoS ONE* 4:e6100. doi: 10.1371/journal.pone.0006100

Celik, N., Webb, C. T., Leyton, D. L., Holt, K. E., Heinz, E., Gorrell, R., et al. (2012). A bioinformatic strategy for the detection, classification and analysis of bacterial autotransporters. *PLoS ONE* 7:e43245. doi: 10.1371/journal.pone.0043245

Chagnot, C., Zorgani, M. A., Astruc, T., and Desvaux, M. (2013). Proteinaceous determinants of surface colonization in bacteria: bacterial adhesion and biofilm formation from a protein secretion perspective. *Front. Microbiol.* 4:303. doi: 10.3389/fmicb.2013.00303

Desvaux, M., Cooper, L. M., Filenko, N. A., Scott-Tucker, A., Turner, S. M., Cole, J. A., et al. (2006). The unusual extended signal peptide region of the Type V secretion system is phylogenetically restricted. *FEMS Microbiol. Lett.* 264, 22–30. doi: 10.1111/j.1574-6968.2006.00425.x

Desvaux, M., Hébraud, M., Talon, R., and Henderson, I. R. (2009). Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. *Trends Microbiol.* 17, 139–145. doi: 10.1016/j.tim.2009.01.004

Desvaux, M., Parham, N. J., and Henderson, I. R. (2003). Le système de sécrétion de type V chez les bactéries Gram-négatives (Type V secretion system in Gram-negative bacteria). *Biofutur* 237, 34–37.

Desvaux, M., Parham, N. J., and Henderson, I. R. (2004a). The autotransporter secretion system. *Res. Microbiol.* 155, 53–60. doi: 10.1016/j.resmic.2003.10.002

Desvaux, M., Parham, N. J., Scott-Tucker, A., and Henderson, I. R. (2004b). The general secretory pathway: a general misnomer? *Trends Microbiol.* 12, 306–309. doi: 10.1016/j.tim.2004.05.002

Desvaux, M., Scott-Tucker, A., Turner, S. M., Cooper, L. M., Huber, D., Nataro, J. P., et al. (2007). A conserved extended signal peptide region directs posttranslational protein translocation via a novel mechanism. *Microbiology* 153, 59–70. doi: 10.1099/mic.0.29091-0

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J. M., et al. (2011). T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* 39, W13–W17. doi: 10.1093/nar/gkr245

Domingo Meza-Aguilar, J., Fromme, P., Torres-Larios, A., Mendoza-Hernández, G., Hernandez-Chiñas, U., Arreguin-Espinosa De Los Monteros, R. A., et al. (2014). X-ray crystal structure of the passenger domain of plasmid encoded toxin (Pet), an autotransporter enterotoxin from enteroaggregative *Escherichia coli* (EAEC). *Biochem. Biophys. Res. Commun.* 445, 439–444. doi: 10.1016/j.bbrc.2014.02.016

Drobnak, I., Braselmann, E., Chaney, J. L., Leyton, D. L., Bernstein, H. D., Lithgow, T., et al. (2015). Of linkers and autochaperones: an unambiguous nomenclature to identify common and uncommon themes for autotransporter secretion. *Mol. Microbiol.* 95, 1–16. doi: 10.1111/mmi.12838

Dutta, P. R., Sui, B. Q., and Nataro, J. P. (2003). Structure-function analysis of the enteroaggregative *Escherichia coli* plasmid-encoded toxin autotransporter using scanning linker mutagenesis. *J. Biol. Chem.* 278, 39912–39920. doi: 10.1074/jbc.M303595200

Economou, A., Christie, P. J., Fernandez, R. C., Palmer, T., Plano, G. V., and Pugsley, A. P. (2006). Secretion by numbers: protein traffic in prokaryotes. *Mol. Microbiol.* 62, 308–319. doi: 10.1111/j.1365-2958.2006.05377.x

Eicher, S. C., and Dehio, C. (2012). Bartonella entry mechanisms into mammalian host cells. *Cell. Microbiol.* 14, 1166–1173. doi: 10.1111/j.1462-5822.2012.01806.x

Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., et al. (2017). InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* 45, D190–D199. doi: 10.1093/nar/gkw1107

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–285. doi: 10.1093/nar/gkv1344

Gangwer, K. A., Mushrush, D. J., Stauff, D. L., Spiller, B., Mcclain, M. S., Cover, T. L., et al. (2007). Crystal structure of the *Helicobacter pylori* vacuolating toxin p55 domain. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16293–16298. doi: 10.1073/pnas.0707447104

Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* 14, 685–695. doi: 10.1093/oxfordjournals.molbev.a025808

Gouet, P., Robert, X., and Courcelle, E. (2003). ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* 31, 3320–3323. doi: 10.1093/nar/gkg556

Haft, D. H., Selengut, J. D., and White, O. (2003). The TIGRFAMs database of protein families. *Nucleic Acids Res.* 31, 371–373. doi: 10.1093/nar/gkg128

Henderson, I. R., and Desvaux, M. (2004). Type V secretion pathway: a premium source source of virulence factors? *Drug Discov. Today* 9, 241.

Henderson, I. R., Navarro-Garcia, F., Desvaux, M., Fernandez, R. C., and Ala'aldeen, D. (2004). Type V protein secretion pathway: the autotransporter story. *Microbiol. Mol. Biol. Rev.* 68, 692–744. doi: 10.1128/MMBR.68.4.692-744.2004

Heras, B., Totsika, M., Peters, K. M., Paxman, J. J., Gee, C. L., Jarrott, R. J., et al. (2014). The antigen 43 structure reveals a molecular Velcro-like mechanism of autotransporter-mediated bacterial clumping. *Proc. Natl. Acad. Sci. U.S.A.* 111, 457–462. doi: 10.1073/pnas.1311592111

Jenkins, J., Mayans, O., and Pickersgill, R. (1998). Structure and evolution of parallel β-helix proteins. *J. Struct. Biol.* 122, 236–246. doi: 10.1006/jsbi.1998.3985

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., Mcanulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Jong, W. S., and Luirink, J. (2008). The conserved extension of the Hbp autotransporter signal peptide does not determine targeting pathway specificity. *Biochem. Biophys. Res. Commun.* 368, 522–527. doi: 10.1016/j.bbrc.2008.01.122

Junker, M., Schuster, C. C., Mcdonnell, A. V., Sorg, K. A., Finn, M. C., Berger, B., et al. (2006). Pertactin β-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. *Proc. Natl. Acad. Sci. U.S.A.* 103, 4918–4923. doi: 10.1073/pnas.0507923103

Kajava, A. V., Cheng, N., Cleaver, R., Kessel, M., Simon, M. N., Willery, E., et al. (2001). Beta-helix model for the filamentous haemagglutinin adhesin of Bordetella pertussis and related bacterial secretory proteins. *Mol. Microbiol.* 42, 279–292. doi: 10.1046/j.1365-2958.2001.02598.x

Kajava, A. V., and Steven, A. C. (2006). The turn of the screw: variations of the abundant β-solenoid motif in passenger domains of Type V secretory proteins. *J. Struct. Biol.* 155, 306–315. doi: 10.1016/j.jsb.2006.01.015

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053

Kelley, L. A., and Sternberg, M. J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* 4, 363–371. doi: 10.1038/nprot.2009.2

Khan, S., Mian, H. S., Sandercock, L. E., Chirgadze, N. Y., and Pai, E. F. (2011). Crystal structure of the passenger domain of the *Escherichia coli* autotransporter EspP. *J. Mol. Biol.* 413, 985–1000. doi: 10.1016/j.jmb.2011.09.028

Kloepper, T. H., and Huson, D. H. (2008). Drawing explicit phylogenetic networks and their integration into SplitsTree. *BMC Evol. Biol.* 8:22. doi: 10.1186/1471-2148-8-22

Kühnel, K., and Diezmann, D. (2011). Crystal structure of the autochaperone region from the *Shigella flexneri* autotransporter IcsA. *J. Bacteriol.* 193, 2042–2045. doi: 10.1128/JB.00790-10

Lam, S. D., Dawson, N. L., Das, S., Sillitoe, I., Ashford, P., Lee, D., et al. (2016). Gene3D: expanding the utility of domain assignments. *Nucleic Acids Res.* 44, D404–D409. doi: 10.1093/nar/gkv1231

Leo, J. C., Grin, I., and Linke, D. (2012). Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. *Philos. Trans. R. Soc. B. Biol. Sci.* 367, 1088–1101. doi: 10.1098/rstb.2011.0208

Leyton, D. L., Rossiter, A. E., and Henderson, I. R. (2012). From self sufficiency to dependence: mechanisms and factors important for autotransporter biogenesis. *Nat. Rev. Microbiol.* 10, 213–225. doi: 10.1038/nrmicro2733

Magis, C., Taly, J. F., Bussotti, G., Chang, J. M., Di Tommaso, P., Erb, I., et al. (2014). T-Coffee: tree-based consistency objective function for alignment evaluation. *Methods Mol. Biol.* 1079, 117–129. doi: 10.1007/978-1-62703-646-7_7

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., et al. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 45, D200–D203. doi: 10.1093/nar/gkw1129

May, K. L., and Morona, R. (2008). Mutagenesis of the *Shigella flexneri* autotransporter IcsA reveals novel functional regions involved in IcsA biogenesis and recruitment of host neural Wiscott-Aldrich syndrome protein. *J. Bacteriol.* 190, 4666–4676. doi: 10.1128/JB.00093-08

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., et al. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* 45, D183–D189. doi: 10.1093/nar/gkw1138

Nicolay, T., Vanderleyden, J., and Spaepen, S. (2015). Autotransporter-based cell surface display in Gram-negative bacteria. *Crit. Rev. Microbiol.* 41, 109–123. doi: 10.3109/1040841X.2013.804032

Notredame, C., Higgins, D. G., and Heringa, J. (2000). T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302, 205–217. doi: 10.1006/jmbi.2000.4042

O'sullivan, O., Suhre, K., Abergel, C., Higgins, D. G., and Notredame, C. (2004). 3DCoffee: combining protein sequences and structures within multiple sequence alignments. *J. Mol. Biol.* 340, 385–395. doi: 10.1016/j.jmb.2004.04.058

Ohnishi, Y., and Horinouchi, S. (1996). Extracellular production of a *Serratia marcescens* serine protease in *Escherichia coli*. *Biosci. Biotechnol. Biochem.* 60, 1551–1558. doi: 10.1271/bbb.60.1551

Ohnishi, Y., Nishiyama, M., Horinouchi, S., and Beppu, T. (1994). Involvement of the COOH-terminal pro-sequence of *Serratia marcescens* serine protease in the folding of the mature enzyme. *J. Biol. Chem.* 269, 32800–32806.

Oliver, D. C., Huang, G., Nodel, E., Pleasance, S., and Fernandez, R. C. (2003). A conserved region within the *Bordetella pertussis* autotransporter BrkA is necessary for folding of its passenger domain. *Mol. Microbiol.* 47, 1367–1383. doi: 10.1046/j.1365-2958.2003.03377.x

Oomen, C. J., Van Ulsen, P., Van Gelder, P., Feijen, M., Tommassen, J., and Gros, P. (2004). Structure of the translocator domain of a bacterial autotransporter. *EMBO J.* 23, 1257–1266. doi: 10.1038/sj.emboj.7600148

Otto, B. R., Sijbrandi, R., Luirink, J., Oudega, B., Heddle, J. G., Mizutani, K., et al. (2005). Crystal structure of hemoglobin protease, a heme binding autotransporter protein from pathogenic *Escherichia coli*. *J. Biol. Chem.* 280, 17339–17345. doi: 10.1074/jbc.M412885200

Pedruzzi, I., Rivoire, C., Auchincloss, A. H., Coudert, E., Keller, G., De Castro, E., et al. (2015). HAMAP in 2015: updates to the protein family classification and annotation system. *Nucleic Acids Res.* 43, D1064–D1070. doi: 10.1093/nar/gku1002

Potenza, E., Di Domenico, T., Walsh, I., and Tosatto, S. C. (2015). MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins. *Nucleic Acids Res.* 43, D315–D320. doi: 10.1093/nar/gku982

Renn, J. P., Junker, M., Besingi, R. N., Braselmann, E., and Clark, P. L. (2012). ATP-independent control of autotransporter virulence protein transport via the folding properties of the secreted protein. *Chem. Biol.* 19, 287–296. doi: 10.1016/j.chembiol.2011.11.009

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320–W324. doi: 10.1093/nar/gku316

Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. U.S.A.* 95, 5857–5864. doi: 10.1073/pnas.95.11.5857

Selkrig, J., Mosbahi, K., Webb, C. T., Belousoff, M. J., Perry, A. J., Wells, T. J., et al. (2012). Discovery of an archetypal protein transport system in bacterial outer membranes. *Nat. Struct. Mol. Biol.* 19, 506–510, S501. doi: 10.1038/nsmb.2261

Servant, F., Bru, C., Carrère, S., Courcelle, E., Gouzy, J., Peyruc, D., et al. (2002). ProDom: automated clustering of homologous domains. *Brief. Bioinform.* 3, 246–251. doi: 10.1093/bib/3.3.246

Sigrist, C. J., De Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., et al. (2013). New and continuing developments at PROSITE. *Nucleic Acids Res.* 41, D344–D347. doi: 10.1093/nar/gks1067

Sillitoe, I., Lewis, T. E., Cuff, A., Das, S., Ashford, P., Dawson, N. L., et al. (2015). CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res.* 43, D376–D381. doi: 10.1093/nar/gku947

Skillman, K. M., Barnard, T. J., Peterson, J. H., Ghirlando, R., and Bernstein, H. D. (2005). Efficient secretion of a folded protein domain by a monomeric bacterial autotransporter. *Mol. Microbiol.* 58, 945–958. doi: 10.1111/j.1365-2958.2005.04885.x

Soprova, Z., Sauri, A., Van Ulsen, P., Tame, J. R., Den Blaauwen, T., Jong, W. S., et al. (2010). A conserved aromatic residue in the autochaperone domain of the autotransporter Hbp is critical for initiation of outer membrane translocation. *J. Biol. Chem.* 285, 38224–38233. doi: 10.1074/jbc.M110.180505

Sutcliffe, I. C. (2010). A phylum level perspective on bacterial cell envelope architecture. *Trends Microbiol.* 18, 464–470. doi: 10.1016/j.tim.2010.06.005

Van Den Berg, B. (2010). Crystal structure of a full-length autotransporter. *J. Mol. Biol.* 396, 627–633. doi: 10.1016/j.jmb.2009.12.061

Van Ulsen, P., Rahman, S., Jong, W. S., Daleke-Schermerhorn, M. H., and Luirink, J. (2014). Type V secretion: from biogenesis to biotechnology. *Biochim. Biophys. Acta* 1843, 1592–1611. doi: 10.1016/j.bbamcr.2013.11.006

Velarde, J. J., and Nataro, J. P. (2004). Hydrophobic residues of the autotransporter EspP linker domain are important for outer membrane translocation of its passenger. *J. Biol. Chem.* 279, 31495–31504. doi: 10.1074/jbc.M404424200

Vo, J. L., Martinez Ortiz, G. C., Subedi, P., Keerthikumar, S., Mathivanan, S., Paxman, J. J., et al. (2017). Autotransporter adhesins in *Escherichia coli* pathogenesis. *Proteomics* doi: 10.1002/pmic.201600431

Wells, T. J., Tree, J. J., Ulett, G. C., and Schembri, M. A. (2007). Autotransporter proteins: novel targets at the bacterial cell surface. *FEMS Microbiol. Lett.* 274, 163–172. doi: 10.1111/j.1574-6968.2007.00833.x

Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., et al. (2009). SUPERFAMILY–sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Res.* 37, D380–D386. doi: 10.1093/nar/gkn762

Wu, C. H., Nikolskaya, A., Huang, H., Yeh, L. S., Natale, D. A., Vinayaka, C. R., et al. (2004). PIRSF: family classification system at the Protein Information Resource. *Nucleic Acids Res.* 32, D112–D114. doi: 10.1093/nar/gkh097

Conflict of Interest Statement: RR and MP are permanent employees of GSK, which provided support in the form of salaries.

The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Rojas-Lopez, Zorgani, Kelley, Bailly, Kajava, Henderson, Polticelli, Pizza, Rosini and Desvaux. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Conclusions and Perspectives

Conclusions and Perspectives

By implementing the reverse vaccinology approach to the EHEC O157:H7 genome, combined with the GMMA delivery system, we have identified new potential vaccine candidates. *In silico* analysis of an EHEC genome prototype allowed us to go beyond known antigens, as LEE genes, hitherto employed in vaccine development for EHEC. What is more, gene variability and distribution analysis allow us to select prevalent vaccine candidates not only in EHEC, but also in other pathotypes as EPEC. In particular, immunization with MC001 using the GMMA delivery system reduced intestinal EHEC O157:H7 colonization and lowered bacterial count in feces, colon and cecum in mice. MC001 was found to be homologous to the *Salmonella* Typhimurium lipid A deacylase enzyme (LpxR) and to our knowledge this study is the first to describe it as vaccine candidate.

Reverse vaccinology has become a useful tool in vaccine discovery and development by reducing the time needed to find new antigens. Whole genome analysis, like the prediction of the functions and subcellular localizations of coding DNA sequences, is the core of reverse vaccinology. During this work, we performed in-depth bioinformatic analysis of protein features and domains. However, for some of them we were not able to predict specific features that would let us classify them as potential antigens (*i.e.* exporter or secreted proteins). Furthermore, errors in the genome annotations lead to the exclusion of other potential antigens. Following on from this study, it will be of great interest to analyze further the properties of the 24 selected antigens. In our findings, we selected some proteins that are still unknown or whose properties are not fully elucidated, so there is a need for extensive functional characterization and elucidation of their role in pathogenesis. In addition to all the above considerations, the inclusion of RNA-seq data provided supplementary information about the expression of antigens in certain conditions. Recently, with reverse vaccinology 2.0 (Rappuoli et al., 2016), the new findings of protective human antibody features, such as structure and epitopes, are refining the discovery of new vaccines. Furthermore, the inclusion of other data from omics approaches such as proteomics, and transcriptomics data, will help to target new antigens at the right moment of protein expression. Likewise, for vaccine development, the immune properties of the 21 prospective antigens identified in this work remain to be studied, using *in vitro* and *in vivo* models.

Further studies are required to improve antigen production and expression and more specifically for vaccine development purposes. It is well known that a protein used as a vaccine

must be presented as in its native form showing suitable epitopes in order to trigger a specific immune response against these antigens. Antigens purified as soluble recombinant proteins will help production of native forms of the antigens in question. During this work, some of the antigens selected, including MC001, were purified as insoluble proteins, even after standardizing expression and purification. Moreover, it has been shown that additional technologies can facilitate appropriate delivery of antigens. For this reason, and as shown in previous studies, combining insoluble antigens with the GMMA system (as MC001) is one of the main goals in improving epitope presentation.

Specific questions about MC001 need to be resolved, as, for example, whether the effect of GMMA-MC001 is directly attributed to the MC001/LpxR protein or whether it is attributed to the possible changes this protein can make to LPS, because of its deacylase properties. Moreover, the characterization of MC001/LpxR in EHEC will be required to elucidate its role in the pathogenesis of these bacteria. Such analyses can include deacylation studies of LPS, to uncover how it affects EHEC virulence or symbiosis in the host and whether it is capable of regulating the immune response like their homologues in *H. pylori*. Furthermore, it will be necessary to gain more insight into how the immune response can be triggered by a vaccine, by testing different routes of immunization that could enhance the immune response and protection. Vaccine preparation will also require studies of large-scale production and formulation, including different adjuvants to identify the best for further vaccine testing in human subjects. We have tested standard laboratory conditions of expression that can be the basis for further process development and also for standard formulation procedures currently used for human vaccination.

We found that MC001 afforded protection when delivered by GMMA, and more studies are needed to characterize in detail the other promising antigens tested in this study (MC007 and MC020). It is known that autotransporters, like MC020, are able to trigger the immune response and work well as vaccine candidates. Our results showed that MC020 raised specific IgG antibodies, but did not confer immune protection after challenge with EHEC. For these reasons, new studies on this antigen, such as modifying its expression or engineering its different forms, could lead to increased protection against EHEC colonization.

Reverse vaccinology targets exposed antigens, such as outer membrane or exported proteins. A clear example of such proteins are the autotransporters, which are widely employed in vaccine development for other InPEC, as ETEC. The aim of the second part of this thesis was to analyze these structures, mainly to reveal the conformation and exposure of prospective epitopes, so as to shed light on the immunological trigger response and protection against the

pathogens expressing these structures. Autochaperones, associated with numerous β-helical passengers, are important for the folding and secretion of autotransporters, and so are a good target for different applications. For the biotechnological applications, AC can be used for the efficient expression of heterologous proteins for cell surface display via T5aSS. For biomedical applications, AC can be used to block AT secretion, therefore reducing the virulence of pathogenic bacteria. AC could be used as prospective target epitopes for vaccine development. Moreover, the GMMA system has also proven suitable for delivery of outer membrane antigens and if, for instance, autochaperones play an important role in the assembly of autotransporters, it will be very interesting to target the autochaperones combined with the GMMA system, so they can be presented in natural conditions that mimic those present in the bacteria during infection, thereby improving protection.

This thesis has gone into some detail about the use of the GMMA system to improve vaccines. Forthcoming tasks include characterization of the GMMA system for InPEC vaccines. There is a need to characterize the proteome of outer membrane vesicles from both GMMA and NOMV, and particularly to investigate NOMV from EHEC, work that has already been initiated by other research groups. For instance, information on the role of NOMV in biofilm formation, attachment, or evasion of the immune response, among other features of bacterial pathogenesis (Kulp *et al.*, 2010; Schwechheimer *et al.,* 2015), and on the proteins involved in these processes might also be used in the search for new vaccines.

As a corollary, the highly genomic plasticity of *E. coli* is also one of the hurdles to be overcome in developing an effective treatment against the diseases caused by this bacterium. *E. coli* strains share a highly conserved common core genome, while the encoding genes related to diverse pathogenic functions involved in infection, disease development, and host-pathogen interaction vary among pathotypes. So, a universal vaccine candidate has not yet been identified. The inclusion of a combination of antigens might cover infections caused not only by EHEC, but by other InPEC too.

Efforts to prevent diarrheagenic diseases produced by intestinal pathogenic *E. coli* should be strengthened, not only for the most common InPEC, as EHEC or ETEC, but also for other pathotypes that are becoming relevant to public health. Treatments and preventive interventions should target non-O157 EHEC serotypes, and other InPEC pathotypes, as EPEC, DAEC, EIEC, and EAEC, for which vaccine development and discovery studies are scarce or nonexistent (Bouzari *et al.*, 2010). Besides the canonical serotype O157:H7, EHEC infections and associated diseases can be related to 6 other STEC serotypes, O26:H11, O45:H2, O103:H11, O111, O121:H19, O145. Nowadays, it has been proposed that multidisciplinary research,

collaborations, and partnerships should integrate the One Health concept, as done in Latin America in recent years (Torres, 2017). In particular, this concept states that human health is connected to the relationship between the environment, animals, and human beings. For this reason, research should be extended to animal pathogenic strains. Examples of this are *E. coli* research studies in animal reservoirs (Etcheverría *et al.*, 2016). Overall, the battle against InPEC infections is ongoing and the perspectives for vaccine development need a multidisciplinary approach to make advances in treatment and prevention. Additional efforts, not only in biomedical or basic research, but also in the clinical, veterinarian, and public health fields, will be required to shed light on disease proliferation in different settings and to design new preventive strategies (CDC). The data generated by such studies will be important for the development of new preventive leads as vaccines against this relevant human and animal pathogen.

References

Introduction

Bell, B.P., Goldoft, M., Griffin, P.M., Davis, M.A., Gordon, D.C., Tarr, P.I., *et al.* (1994). A multistate outbreak of Escherichia coli O157:H7-associated bloody diarrhea and hemolytic uremic syndrome from hamburgers. The Washington experience. *JAMA* 272(17)**,** 1349-1353.

Boerlin, P., McEwen, S.A., Boerlin-Petzold, F., Wilson, J.B., Johnson, R.P., and Gyles, C.L. (1999). Associations between virulence factors of Shiga toxin-producing Escherichia coli and disease in humans. *J Clin Microbiol* 37(3)**,** 497-503.

Bouzari, S., Dashti, A., Jafari, A., and Oloomi, M. (2010). Immune response against adhesins of enteroaggregative Escherichia coli immunized by three different vaccination strategies (DNA/DNA, Protein/Protein, and DNA/Protein) in mice. *Comp Immunol Microbiol Infect Dis* 33**,** 215-225. 10.1016/j.cimid.2008.10.002

Burland, V., Shao, Y., Perna, N.T., Plunkett, G., Sofia, H.J., and Blattner, F.R. (1998). The complete DNA sequence and analysis of the large virulence plasmid of Escherichia coli O157:H7. *Nucleic Acids Res* 26(18)**,** 4196-4204.

Caprioli, A., Morabito, S., Brugere, H., and Oswald, E. (2005). Enterohaemorrhagic Escherichia coli: emerging issues on virulence and modes of transmission. *Vet Res* 36(3)**,** 289-311. doi: 10.1051/vetres:2005002.

CDC Available: http://www.cdc.gov/onehealth/basics/index.html [Accessed].

Coombes, B.K., Gilmour, M.W., and Goodman, C.D. (2011). The evolution of virulence in non-o157 shiga toxin-producing Escherichia coli. *Front Microbiol* 2**,** 90. doi: 10.3389/fmicb.2011.00090.

Donnenberg, M.S., and Whittam, T.S. (2001). Pathogenesis and evolution of virulence in enteropathogenic and enterohemorrhagic Escherichia coli. *J Clin Invest* 107(5)**,** 539-548. doi: 10.1172/JCI12404.

Doughty, S., Sloan, J., Bennett-Wood, V., Robertson, M., Robins-Browne, R.M., and Hartland, E.L. (2002). Identification of a novel fimbrial gene cluster related to long polar fimbriae in locus of enterocyte effacement-negative strains of enterohemorrhagic Escherichia coli. *Infect Immun* 70(12)**,** 6761-6769.

Etcheverría, A.I., Lucchesi, P.M.A., Krüger, A., Bentancor, A.B., and Padola, N.L. (2016). "Escherichia coli in Animals," in *Escherichia coli in the Americas,* ed. A.G. Torres. (Cham: Springer International Publishing), 149-172.

Herold, S., Paton, J.C., and Paton, A.W. (2009). Sab, a novel autotransporter of locus of enterocyte effacement-negative shiga-toxigenic Escherichia coli O113:H21, contributes to adherence and biofilm formation. *Infect Immun* 77(8), 3234-3243. doi: 10.1128/IAI.00031-09.

Holtz, L.R., Neill, M.A., and Tarr, P.I. (2009). Acute bloody diarrhea: a medical emergency for patients of all ages. *Gastroenterology* 136(6), 1887-1898. doi: 10.1053/j.gastro.2009.02.059.

Kaper, J.B., Nataro, J.P., and Mobley, H.L. (2004). Pathogenic Escherichia coli. *Nat Rev Microbiol* 2(2), 123-140. doi: 10.1038/nrmicro818.

Karmali, M.A., Mascarenhas, M., Shen, S., Ziebell, K., Johnson, S., Reid-Smith, R., *et al.* (2003). Association of genomic O island 122 of Escherichia coli EDL 933 with verocytotoxin-producing Escherichia coli seropathotypes that are linked to epidemic and/or serious disease. *J Clin Microbiol* 41(11), 4930-4940.

Kimmitt, P.T., Harwood, C.R., and Barer, M.R. (2000). Toxin gene expression by shiga toxin-producing Escherichia coli: the role of antibiotics and the bacterial SOS response. *Emerg Infect Dis* 6(5), 458-465. doi: 10.3201/eid0605.000503.

Lewis, S.B., Cook, V., Tighe, R., and Schuller, S. (2015). Enterohemorrhagic Escherichia coli colonization of human colonic epithelium in vitro and ex vivo. *Infect Immun* 83(3), 942-949. doi: 10.1128/IAI.02928-14.

Lu, Y., Iyoda, S., Satou, H., Satou, H., Itoh, K., Saitoh, T., *et al.* (2006). A new immunoglobulin-binding protein, EibG, is responsible for the chain-like adhesion phenotype of locus of enterocyte effacement-negative, shiga toxin-producing Escherichia coli. *Infect Immun* 74(10), 5747-5755. doi: 10.1128/IAI.00724-06.

McGannon, C.M., Fuller, C.A., and Weiss, A.A. (2010). Different classes of antibiotics differentially influence shiga toxin production. *Antimicrob Agents Chemother* 54(9), 3790-3798. doi: 10.1128/AAC.01783-09.

McWilliams, B.D., and Torres, A.G. (2014). Enterohemorrhagic Escherichia coli Adhesins. *Microbiol Spectr* 2(3). doi: 10.1128/microbiolspec.EHEC-0003-2013.

Mejias, M.P., Hiriart, Y., Lauche, C., Fernandez-Brando, R.J., Pardo, R., Bruballa, A., *et al.* (2016). Development of camelid single chain antibodies against Shiga toxin type 2 (Stx2) with therapeutic potential against Hemolytic Uremic Syndrome (HUS). *Sci Rep* 6, 24913. doi: 10.1038/srep24913.

Merkel, V., Ohder, B., Bielaszewska, M., Zhang, W., Fruth, A., Menge, C., *et al.* (2010). Distribution and phylogeny of immunoglobulin-binding protein G in Shiga toxin-

producing Escherichia coli and its association with adherence phenotypes. *Infect Immun* 78(8)**,** 3625-3636. doi: 10.1128/IAI.00006-10.

Monteiro, R., Ageorges, V., Rojas-Lopez, M., Schmidt, H., Weiss, A., Bertin, Y., *et al.* (2016). A secretome view of colonisation factors in Shiga toxin-encoding Escherichia coli (STEC): from enterohaemorrhagic E. coli (EHEC) to related enteropathotypes. *FEMS Microbiol Lett* 363(16). doi: 10.1093/femsle/fnw179.

Nataro, J.P., and Kaper, J.B. (1998a). Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev* 11(1)**,** 142-201.

Nataro, J.P., and Kaper, J.B. (1998b). Diarrheogenic *Escherichia coli*. . *Clin Microbiol Rev* 11**,** 142-210

Pacheco, A.R., and Sperandio, V. (2012). Shiga toxin in enterohemorrhagic E.coli: regulation and novel anti-virulence strategies. *Front Cell Infect Microbiol* 2**,** 81. doi: 10.3389/fcimb.2012.00081.

Paton, A.W., Srimanote, P., Woodrow, M.C., and Paton, J.C. (2001). Characterization of Saa, a novel autoagglutinating adhesin produced by locus of enterocyte effacement-negative Shiga-toxigenic Escherichia coli strains that are virulent for humans. *Infect Immun* 69(11)**,** 6999-7009. doi: 10.1128/IAI.69.11.6999-7009.2001.

Perna, N.T., Plunkett, G., Burland, V., Mau, B., Glasner, J.D., Rose, D.J., *et al.* (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409**,** 529-532.

Rasko, D.A., Webster, D.R., Sahl, J.W., Bashir, A., Boisen, N., Scheutz, F., *et al.* (2011). Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N Engl J Med* 365(8)**,** 709-717. doi: 10.1056/NEJMoa1106920.

Rendon, M.A., Saldana, Z., Erdem, A.L., Monteiro-Neto, V., Vazquez, A., Kaper, J.B., *et al.* (2007). Commensal and pathogenic Escherichia coli use a common pilus adherence factor for epithelial cell colonization. *Proc Natl Acad Sci U S A* 104(25)**,** 10637-10642. doi: 0704104104 [pii]

10.1073/pnas.0704104104.

Rivas, M., Chinen, I., and Guth, B.E.C. (2016). "Enterohemorrhagic (Shiga Toxin-Producing) Escherichia coli," in *Escherichia coli in the Americas,* ed. A.G. Torres. (Cham: Springer International Publishing), 97-123.

Samadder, P., Xicohtencatl-Cortes, J., Saldana, Z., Jordan, D., Tarr, P.I., Kaper, J.B., *et al.* (2009). The Escherichia coli ycbQRST operon encodes fimbriae with laminin-binding and epithelial cell adherence properties in Shiga-toxigenic E. coli O157:H7. *Environ Microbiol*. doi: EMI1906 [pii]

10.1111/j.1462-2920.2009.01906.x.

Scheutz, F., Teel, L.D., Beutin, L., Pierard, D., Buvens, G., Karch, H., *et al.* (2012). Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. *J Clin Microbiol* 50(9)**,** 2951-2963. doi: 10.1128/JCM.00860-12.

Stacy M. Crim, P.M.G., Robert Tauxe, Ellyn P. Marder, Debra Gilliss, Alicia B. Cronquist, , and Matthew Cartter, M.T.-D.A., David Blythe, Kirk Smith, Sarah Lathrop, Shelley Zansky, Paul R. Cieslak, John Dunn, Kristin G. Holt, Beverly Wolpert, Olga L. Henao. (2015). "Preliminary Incidence and Trends of Infection with Pathogens Transmitted Commonly Through Food — Foodborne Diseases Active Surveillance Network, 10 U.S. Sites, 2006–2014", in: *MMWR Morb Mortal Wkly Rep.* (ed.) M.a.M.W.R. (MMWR). (Atlanta: Center for Surveillance, Epidemiology, and Laboratory Services, Centers for Disease Control and Prevention (CDC)).

Stevens M.P., Roe A.J., Vlisidou I., van Diemen P.M., La Ragione R.M., Best A., Woodward M.J., Gally D.L., Wallis T.S. 2004. Mutation of toxB and a truncated version of the efa-1 gene in Escherichia coli O157:H7 influences the expression and secretion of locus of enterocyte effacement-encoded proteins but not intestinal colonization in calves or sheep. Infect. Immun. 72:5402–5411.

Stevens, M.P.T.S.W. (2005). "Adhesins of Enterohemorrhagic Escherichia coli.," in *E. coli and Salmonella: cellular and molecular biology* ed. M.K.B.K.B. Low. (Washington.).

Tatsuno, I., Horie, M., Abe, H., Miki, T., Makino, K., Shinagawa, H., Taguchi, H., Kamiya, S., Hayashi, T. Sasakawa, C.. (2001). "toxB gene on pO157 of enterohemorrhagic Escherichia coli O157:H7 is required for full epithelial cell adherence phenotype." Infect Immun **69**(11): 6660-6669.

Tarr, P.I., Gordon, C.A., and Chandler, W.L. (2005). Shiga-toxin-producing Escherichia coli and haemolytic uraemic syndrome. *Lancet* 365(9464)**,** 1073-1086. doi: 10.1016/S0140-6736(05)71144-2.

Thorpe, C.M. (2004). Shiga toxin-producing Escherichia coli infection. *Clin Infect Dis* 38(9)**,** 1298-1303. doi: 10.1086/383473

CID32684 [pii].

Torres, A.G. (2008). "Intestinal Pathogenic Escherichia coli," in *Vaccines for Biodefense and Emerging and Neglected Diseases,* ed. L.S. Alan D.T. Barrett. Elsevier Inc.), 1003-1020.

Torres, A.G., Giron, J.A., Perna, N.T., Burland, V., Blattner, F.R., Avelino-Flores, F., *et al.* (2002). Identification and Characterization of *lpfABCC'DE*, a Fimbrial Operon of Enterohemorrhagic *Escherichia coli* O157:H7. *Infect Immun* 70**,** 5416–5427.

Torres, A.G., Kanack, K.J., Tutt, C.B., Popov, V., and Kaper, J.B. (2004). Characterization of the second long polar (LP) fimbriae of Escherichia coli O157:H7 and distribution of LP fimbriae in other pathogenic E. coli strains. *FEMS Microbiol Lett* 238(2)**,** 333-344. doi: 10.1016/j.femsle.2004.07.053 S0378-1097(04)00570-1 [pii].

Torres, A.G. (2017). Escherichia coli diseases in Latin America-a 'One Health' multidisciplinary approach. *Pathog Dis* 75. 10.1093/femspd/ftx012

Torres, A.G., Zhou, X., and Kaper, J.B. (2005). Adherence of Diarrheagenic *Escherichia coli* Strains to Epithelial cells. *Infect Immun* 73**,** 18–29.

Xicohtencatl-Cortes, J., Monteiro-Neto, V., Ledesma, M.A., Jordan, D.M., Francetic, O., Kaper, J.B., *et al.* (2007). Intestinal adherence associated with type IV pili of enterohemorrhagic Escherichia coli O157:H7. *J Clin Invest* 117(11)**,** 3519-3529. doi: 10.1172/JCI30727.

Xicohtencatl-Cortes, J., Monteiro-Neto, V., Saldana, Z., Ledesma, M.A., Puente, J.L., and Giron, J.A. (2009). The type 4 pili of enterohemorrhagic Escherichia coli O157:H7 are multipurpose structures with pathogenic attributes. *J Bacteriol* 191(1)**,** 411-421. doi: JB.01306-08 [pii]

10.1128/JB.01306-08.

Zhang, X., McDaniel, A.D., Wolf, L.E., Keusch, G.T., Waldor, M.K., and Acheson, D.W. (2000). Quinolone antibiotics induce Shiga toxin-encoding bacteriophages, toxin production, and death in mice. *J Infect Dis* 181(2)**,** 664-670. doi: 10.1086/315239.

First Part

Chapter 1

Ahmed, A., Li, J., Shiloach, Y., Robbins, J.B., and Szu, S.C. (2006). Safety and immunogenicity of Escherichia coli O157 O-specific polysaccharide conjugate vaccine in 2-5-year-old children. *J Infect Dis* 193(4)**,** 515-521. doi: 10.1086/499821.

Amani, J., Mousavi, S.L., Rafati, S., and Salmanian, A.H. (2009). In silico analysis of chimeric espA, eae and tir fragments of Escherichia coli O157:H7 for oral immunogenic applications. *Theor Biol Med Model* 6**,** 28. doi: 10.1186/1742-4682-6-28.

Amani, J., Mousavi, S.L., Rafati, S., and Salmanian, A.H. (2011). Immunogenicity of a plant-derived edible chimeric EspA, Intimin and Tir of Escherichia coli O157:H7 in mice. *Plant Sci* 180(4)**,** 620-627. doi: 10.1016/j.plantsci.2011.01.004.

Amani, J., Salmanian, A.H., Rafati, S., and Mousavi, S.L. (2010). Immunogenic properties of chimeric protein from espA, eae and tir genes of Escherichia coli O157:H7. *Vaccine* 28(42), 6923-6929. doi: 10.1016/j.vaccine.2010.07.061.

Bartolini, E., Ianni, E., Frigimelica, E., Petracca, R., Galli, G., Berlanda Scorza, F., *et al.* (2013). Recombinant outer membrane vesicles carrying Chlamydia muridarum HtrA induce antibodies that neutralize chlamydial infection in vitro. *J Extracell Vesicles* 2. doi: 10.3402/jev.v2i0.20181.

Bentancor, L.V., Bilen, M., Brando, R.J., Ramos, M.V., Ferreira, L.C., Ghiringhelli, P.D., *et al.* (2009). A DNA vaccine encoding the enterohemorragic Escherichia coli Shiga-like toxin 2 A2 and B subunits confers protective immunity to Shiga toxin challenge in the murine model. *Clin Vaccine Immunol* 16(5), 712-718. doi: 10.1128/CVI.00328-08.

Berlanda Scorza, F., Colucci, A.M., Maggiore, L., Sanzone, S., Rossi, O., Ferlenghi, I., *et al.* (2012). High yield production process for Shigella outer membrane particles. *PLoS One* 7(6), e35616. doi: 10.1371/journal.pone.0035616.

Bitzan, M., Poole, R., Mehran, M., Sicard, E., Brockus, C., Thuning-Roberson, C., *et al.* (2009). Safety and pharmacokinetics of chimeric anti-Shiga toxin 1 and anti-Shiga toxin 2 monoclonal antibodies in healthy volunteers. *Antimicrob Agents Chemother* 53(7), 3081-3087. doi: 10.1128/AAC.01661-08.

Brzuszkiewicz, E., Thurmer, A., Schuldes, J., Leimbach, A., Liesegang, H., Meyer, F.D., *et al.* (2011). Genome sequence analyses of two isolates from the recent *Escherichia coli* outbreak in Germany reveal the emergence of a new pathotype: Entero-Aggregative-Haemorrhagic *Escherichia coli* (EAHEC). *Arch Microbiol* 193(12), 883-891. doi: 10.1007/s00203-011-0725-6.

Cai, K., Gao, X., Li, T., Hou, X., Wang, Q., Liu, H., *et al.* (2010). Intragastric immunization of mice with enterohemorrhagic Escherichia coli O157:H7 bacterial ghosts reduces mortality and shedding and induces a Th2-type dominated mixed immune response. *Can J Microbiol* 56(5), 389-398. doi: 10.1139/w10-025.

Cai, K., Gao, X., Li, T., Wang, Q., Hou, X., Tu, W., *et al.* (2011). Enhanced immunogenicity of a novel Stx2Am-Stx1B fusion protein in a mice model of enterohemorrhagic Escherichia coli O157:H7 infection. *Vaccine* 29(5), 946-952. doi: 10.1016/j.vaccine.2010.11.035.

Cai, K., Tu, W., Liu, Y., Li, T., and Wang, H. (2015). Novel fusion antigen displayed-bacterial ghosts vaccine candidate against infection of Escherichia coli O157:H7. *Sci Rep* 5, 17479. doi: 10.1038/srep17479.

Calderon Toledo, C., Arvidsson, I., and Karpman, D. (2011). Cross-reactive protection against enterohemorrhagic Escherichia coli infection by enteropathogenic E. coli in a mouse model. *Infect Immun* 79(6)**,** 2224-2233. doi: 10.1128/IAI.01024-10.

Cataldi, A., Yevsa, T., Vilte, D.A., Schulze, K., Castro-Parodi, M., Larzabal, M., *et al.* (2008). Efficient immune responses against Intimin and EspB of enterohaemorragic Escherichia coli after intranasal vaccination using the TLR2/6 agonist MALP-2 as adjuvant. *Vaccine* 26(44)**,** 5662-5667. doi: 10.1016/j.vaccine.2008.07.027.

Cheng, Y., Feng, Y., Luo, P., Gu, J., Yu, S., Zhang, W.J., *et al.* (2009). Fusion expression and immunogenicity of EHEC EspA-Stx2Al protein: implications for the vaccine development. *J Microbiol* 47(4)**,** 498-505. doi: 10.1007/s12275-009-0116-8.

Daleke-Schermerhorn, M.H., Felix, T., Soprova, Z., Ten Hagen-Jongman, C.M., Vikstrom, D., Majlessi, L., *et al.* (2014). Decoration of outer membrane vesicles with multiple antigens by using an autotransporter approach. *Appl Environ Microbiol* 80(18)**,** 5854-5865. doi: 10.1128/AEM.01941-14.

De Gregorio, E., and Rappuoli, R. (2012). Vaccines for the future: learning from human immunology. *Microb Biotechnol* 5(2)**,** 149-155. doi: 10.1111/j.1751-7915.2011.00276.x.

Fantappie, L., de Santis, M., Chiarot, E., Carboni, F., Bensi, G., Jousson, O., *et al.* (2014). Antibody-mediated immunity induced by engineered Escherichia coli OMVs carrying heterologous antigens in their lumen. *J Extracell Vesicles* 3. doi: 10.3402/jev.v3.24015.

Feavers, I.M., and Maiden, M.C.J. (2017). Recent Progress in the Prevention of Serogroup B Meningococcal Disease. *Clin Vaccine Immunol* 24(5). doi: 10.1128/CVI.00566-16.

Foster, M.A., Iqbal, J., Zhang, C., McHenry, R., Cleveland, B.E., Romero-Herazo, Y., *et al.* (2015). Enteropathogenic and enteroaggregative E. coli in stools of children with acute gastroenteritis in Davidson County, Tennessee. *Diagn Microbiol Infect Dis* 83(3)**,** 319-324. doi: 10.1016/j.diagmicrobio.2015.07.016.

Frank, C., Werber, D., Cramer, J.P., Askar, M., Faber, M., an der Heiden, M., *et al.* (2011). Epidemic profile of Shiga-toxin-producing Escherichia coli O104:H4 outbreak in Germany. *N Engl J Med* 365(19)**,** 1771-1780. doi: 10.1056/NEJMoa1106483.

Fujii, J., Naito, M., Yutsudo, T., Matsumoto, S., Heatherly, D.P., Yamada, T., *et al.* (2012). Protection by a recombinant Mycobacterium bovis Bacillus Calmette-Guerin vaccine expressing Shiga toxin 2 B subunit against Shiga toxin-producing Escherichia coli in mice. *Clin Vaccine Immunol* 19(12)**,** 1932-1937. doi: 10.1128/CVI.00473-12.

Gansheroff, L.J., Wachtel, M.R., and O'Brien, A.D. (1999). Decreased adherence of enterohemorrhagic Escherichia coli to HEp-2 cells in the presence of antibodies that recognize the C-terminal region of intimin. *Infect Immun* 67(12)**,** 6409-6417.

Gao, X., Cai, K., Li, T., Wang, Q., Hou, X., Tian, R., *et al.* (2011). Novel fusion protein protects against adherence and toxicity of enterohemorrhagic Escherichia coli O157:H7 in mice. *Vaccine* 29(38)**,** 6656-6663. doi: 10.1016/j.vaccine.2011.06.106.

Gao, X., Cai, K., Shi, J., Liu, H., Hou, X., Tu, W., *et al.* (2009). Immunogenicity of a novel Stx2B-Stx1B fusion protein in a mice model of Enterohemorrhagic Escherichia coli O157:H7 infection. *Vaccine* 27(14)**,** 2070-2076. doi: 10.1016/j.vaccine.2009.01.115.

Garcia-Angulo, V.A., Kalita, A., Kalita, M., Lozano, L., and Torres, A.G. (2014). Comparative genomics and immunoinformatics approach for the identification of vaccine candidates for enterohemorrhagic Escherichia coli O157:H7. *Infect Immun* 82(5)**,** 2016-2026. doi: 10.1128/IAI.01437-13.

Gerke, C., Colucci, A.M., Giannelli, C., Sanzone, S., Vitali, C.G., Sollai, L., *et al.* (2015). Production of a Shigella sonnei Vaccine Based on Generalized Modules for Membrane Antigens (GMMA), 1790GAHB. *PLoS One* 10(8)**,** e0134478. doi: 10.1371/journal.pone.0134478.

Giuliani, M.M., Adu-Bobie, J., Comanducci, M., Arico, B., Savino, S., Santini, L., *et al.* (2006). A universal vaccine for serogroup B meningococcus. *Proc Natl Acad Sci U S A* 103(29)**,** 10834-10839. doi: 10.1073/pnas.0603940103.

Goldwater, P.N., and Bettelheim, K.A. (2012). Treatment of enterohemorrhagic Escherichia coli (EHEC) infection and hemolytic uremic syndrome (HUS). *BMC Med* 10**,** 12. doi: 10.1186/1741-7015-10-12.

Harris, J.A., Roy, K., Woo-Rasberry, V., Hamilton, D.J., Kansal, R., Qadri, F., *et al.* (2011). Directed evaluation of enterotoxigenic Escherichia coli autotransporter proteins as putative vaccine candidates. *PLoS Negl Trop Dis* 5(12)**,** e1428. doi: 10.1371/journal.pntd.0001428.

Kaper, J.B., Nataro, J.P., and Mobley, H.L. (2004). Pathogenic Escherichia coli. *Nat Rev Microbiol* 2(2)**,** 123-140. doi: 10.1038/nrmicro818.

Konadu, E., Donohue-Rolfe, A., Calderwood, S.B., Pozsgay, V., Shiloach, J., Robbins, J.B., *et al.* (1999). Syntheses and immunologic properties of Escherichia coli O157 O-specific polysaccharide and Shiga toxin 1 B subunit conjugates in mice. *Infect Immun* 67(11)**,** 6191-6193.

Konadu, E., Robbins, J.B., Shiloach, J., Bryla, D.A., and Szu, S.C. (1994). Preparation, characterization, and immunological properties in mice of Escherichia coli O157 O-specific polysaccharide-protein conjugate vaccines. *Infect Immun* 62(11)**,** 5048-5054.

Konadu, E.Y., Parke, J.C., Jr., Tran, H.T., Bryla, D.A., Robbins, J.B., and Szu, S.C. (1998). Investigational vaccine for Escherichia coli O157: phase 1 study of O157 O-specific polysaccharide-Pseudomonas aeruginosa recombinant exoprotein A conjugates in adults. *J Infect Dis* 177(2)**,** 383-387.

Kotloff, K.L., Nataro, J.P., Blackwelder, W.C., Nasrin, D., Farag, T.H., Panchalingam, S., *et al.* (2013). Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet* 382(9888)**,** 209-222. doi: 10.1016/S0140-6736(13)60844-2.

Kovacs, S.D., Mullholland, K., Bosch, J., Campbell, H., Forouzanfar, M.H., Khalil, I., *et al.* (2015). Deconstructing the differences: a comparison of GBD 2010 and CHERG's approach to estimating the mortality burden of diarrhea, pneumonia, and their etiologies. *BMC Infect Dis* 15**,** 16. doi: 10.1186/s12879-014-0728-4.

Leitner, D.R., Lichtenegger, S., Temel, P., Zingl, F.G., Ratzberger, D., Roier, S., *et al.* (2015). A combined vaccine approach against Vibrio cholerae and ETEC based on outer membrane vesicles. *Front Microbiol* 6**,** 823. doi: 10.3389/fmicb.2015.00823.

Liu, J., Sun, Y., Feng, S., Zhu, L., Guo, X., and Qi, C. (2009). Towards an attenuated enterohemorrhagic Escherichia coli O157:H7 vaccine characterized by a deleted ler gene and containing apathogenic Shiga toxins. *Vaccine* 27(43)**,** 5929-5935. doi: 10.1016/j.vaccine.2009.07.097.

MacLennan, C.A., and Saul, A. (2014). Vaccines against poverty. *Proc Natl Acad Sci U S A* 111(34)**,** 12307-12312. doi: 10.1073/pnas.1400473111.

Mayr, U.B., Haller, C., Haidinger, W., Atrasheuskaya, A., Bukin, E., Lubitz, W., *et al.* (2005). Bacterial ghosts as an oral vaccine: a single dose of Escherichia coli O157:H7 bacterial ghosts protects mice against lethal challenge. *Infect Immun* 73(8)**,** 4810-4817. doi: 10.1128/IAI.73.8.4810-4817.2005.

Mayr, U.B., Kudela, P., Atrasheuskaya, A., Bukin, E., Ignatyev, G., and Lubitz, W. (2012). Rectal single dose immunization of mice with Escherichia coli O157:H7 bacterial ghosts induces efficient humoral and cellular immune responses and protects against the lethal heterologous challenge. *Microb Biotechnol* 5(2)**,** 283-294. doi: 10.1111/j.1751-7915.2011.00316.x.

Mejias, M.P., Hiriart, Y., Lauche, C., Fernandez-Brando, R.J., Pardo, R., Bruballa, A., *et al.* (2016). Development of camelid single chain antibodies against Shiga toxin type 2 (Stx2) with therapeutic potential against Hemolytic Uremic Syndrome (HUS). *Sci Rep* 6**,** 24913. doi: 10.1038/srep24913.

Montero, D., Orellana, P., Gutierrez, D., Araya, D., Salazar, J.C., Prado, V., *et al.* (2014). Immunoproteomic analysis to identify Shiga toxin-producing Escherichia coli outer membrane proteins expressed during human infection. *Infect Immun* 82(11)**,** 4767-4777. doi: 10.1128/IAI.02030-14.

Mora, M., Veggi, D., Santini, L., Pizza, M., and Rappuoli, R. (2003). Reverse vaccinology. *Drug Discov Today* 8(10)**,** 459-464.

Moriel, D.G., Bertoldi, I., Spagnuolo, A., Marchi, S., Rosini, R., Nesta, B., *et al.* (2010). Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic Escherichia coli. *Proc Natl Acad Sci U S A* 107(20)**,** 9072-9077. doi: 10.1073/pnas.0915077107.

Nesta, B., Spraggon, G., Alteri, C., Moriel, D.G., Rosini, R., Veggi, D., *et al.* (2012). FdeC, a novel broadly conserved *Escherichia coli* adhesin eliciting protection against urinary tract infections. *MBio* 3(2). doi: 10.1128/mBio.00010-12.

Nesta, B., Valeri, M., Spagnuolo, A., Rosini, R., Mora, M., Donato, P., *et al.* (2014). SslE elicits functional antibodies that impair in vitro mucinase activity and in vivo colonization by both intestinal and extraintestinal Escherichia coli strains. *PLoS Pathog* 10(5)**,** e1004124. doi: 10.1371/journal.ppat.1004124.

Nieves, W., Petersen, H., Judy, B.M., Blumentritt, C.A., Russell-Lodrigue, K., Roy, C.J., *et al.* (2014). A Burkholderia pseudomallei outer membrane vesicle vaccine provides protection against lethal sepsis. *Clin Vaccine Immunol* 21(5)**,** 747-754. doi: 10.1128/CVI.00119-14.

Oliveira, A.F., Cardoso, S.A., Almeida, F.B., de Oliveira, L.L., Pitondo-Silva, A., Soares, S.G., *et al.* (2012). Oral immunization with attenuated Salmonella vaccine expressing Escherichia coli O157:H7 intimin gamma triggers both systemic and mucosal humoral immunity in mice. *Microbiol Immunol* 56(8)**,** 513-522. doi: 10.1111/j.1348-0421.2012.00477.x.

Petersen, H., Nieves, W., Russell-Lodrigue, K., Roy, C.J., and Morici, L.A. (2014). Evaluation of a Burkholderia pseudomallei Outer Membrane Vesicle Vaccine in Nonhuman Primates. *Procedia Vaccinol* 8**,** 38-42. doi: 10.1016/j.provac.2014.07.007.

Pianciola, L., D'Astek, B.A., Mazzeo, M., Chinen, I., Masana, M., and Rivas, M. (2016). Genetic features of human and bovine Escherichia coli O157:H7 strains isolated in Argentina. *Int J Med Microbiol*. doi: 10.1016/j.ijmm.2016.02.005.

Pires, S.M., Fischer-Walker, C.L., Lanata, C.F., Devleesschauwer, B., Hall, A.J., Kirk, M.D., *et al.* (2015). Aetiology-Specific Estimates of the Global and Regional Incidence and Mortality of Diarrhoeal Diseases Commonly Transmitted through Food. *PLoS One* 10(12)**,** e0142927. doi: 10.1371/journal.pone.0142927.

Pizza, M., Scarlato, V., Masignani, V., Giuliani, M.M., Arico, B., Comanducci, M., *et al.* (2000). Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 287(5459)**,** 1816-1820.

Plotkin, S. (2014). History of vaccination. *Proc Natl Acad Sci U S A* 111(34)**,** 12283-12287. doi: 10.1073/pnas.1400472111.

Riquelme-Neira, R., Rivera, A., Saez, D., Fernandez, P., Osorio, G., Del Canto, F., *et al.* (2015). Vaccination with DNA Encoding Truncated Enterohemorrhagic Escherichia coli (EHEC) Factor for Adherence-1 Gene (efa-1') Confers Protective Immunity to Mice Infected with E. coli O157:H7. *Front Cell Infect Microbiol* 5**,** 104. doi: 10.3389/fcimb.2015.00104.

Rivas, M., Chinen, I., and Guth, B.E.C. (2016). "Enterohemorrhagic (Shiga Toxin-Producing) Escherichia coli," in *Escherichia coli in the Americas,* ed. A.G. Torres. (Cham: Springer International Publishing), 97-123.

Rossi, O., Caboni, M., Negrea, A., Necchi, F., Alfini, R., Micoli, F., *et al.* (2016). Toll-Like Receptor Activation by Generalized Modules for Membrane Antigens from Lipid A Mutants of Salmonella enterica Serovars Typhimurium and Enteritidis. *Clin Vaccine Immunol* 23(4)**,** 304-314. doi: 10.1128/CVI.00023-16.

Roy, K., Bartels, S., Qadri, F., and Fleckenstein, J.M. (2010). Enterotoxigenic Escherichia coli elicits immune responses to multiple surface proteins. *Infect Immun* 78(7)**,** 3027-3035. doi: 10.1128/IAI.00264-10.

Schild, S., Nelson, E.J., and Camilli, A. (2008). Immunization with Vibrio cholerae outer membrane vesicles induces protective immunity in mice. *Infect Immun* 76(10)**,** 4554-4563. doi: 10.1128/IAI.00532-08.

Sjoling, A., von Mentzer, A., and Svennerholm, A.M. (2015). Implications of enterotoxigenic Escherichia coli genomics for vaccine development. *Expert Rev Vaccines* 14(4)**,** 551-560. doi: 10.1586/14760584.2015.996553.

Tapia, D., Ross, B.N., Kalita, A., Kalita, M., Hatcher, C.L., Muruato, L.A., *et al.* (2016). From In silico Protein Epitope Density Prediction to Testing Escherichia coli O157:H7 Vaccine Candidates in a Murine Model of Colonization. *Front Cell Infect Microbiol* 6**,** 94. doi: 10.3389/fcimb.2016.00094.

Torres, A.G. (2017). Escherichia coli diseases in Latin America-a 'One Health' multidisciplinary approach. *Pathog Dis* 75(2). doi: 10.1093/femspd/ftx012.

Wan, C.S., Zhou, Y., Yu, Y., Peng, L.J., Zhao, W., and Zheng, X.L. (2011). B-cell epitope KT-12 of enterohemorrhagic Escherichia coli O157:H7: a novel peptide vaccine candidate. *Microbiol Immunol* 55(4)**,** 247-253. doi: 10.1111/j.1348-0421.2011.00316.x.

Wen, S.X., Teel, L.D., Judge, N.A., and O'Brien, A.D. (2006). A plant-based oral vaccine to protect against systemic intoxication by Shiga toxin type 2. *Proc Natl Acad Sci U S A* 103(18)**,** 7082-7087. doi: 10.1073/pnas.0510843103.

Whitfield, C. (2006). Biosynthesis and assembly of capsular polysaccharides in Escherichia coli. *Annu Rev Biochem* 75**,** 39-68. doi: 10.1146/annurev.biochem.75.103004.142545.

Chapter 2

Adu-Bobie, J., Frankel, G., Bain, C., Goncalves, A.G., Trabulsi, L.R., Douce, G., *et al.* (1998). Detection of intimins alpha, beta, gamma, and delta, four intimin derivatives expressed by attaching and effacing microbial pathogens. *J Clin Microbiol* 36(3)**,** 662-668.

Altegoer, F., and Bange, G. (2015). Undiscovered regions on the molecular landscape of flagellar assembly. *Curr Opin Microbiol* 28**,** 98-105. doi: 10.1016/j.mib.2015.08.011.

Alvarez-Martinez, C.E., and Christie, P.J. (2009). Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev* 73(4)**,** 775-808. doi: 10.1128/MMBR.00023-09.

Antelmann, H., Tjalsma, H., Voigt, B., Ohlmeier, S., Bron, S., van Dijl, J.M., *et al.* (2001). A proteomic view on genome-based signal peptide predictions. *Genome Res* 11(9)**,** 1484-1502. doi: 10.1101/gr.182801.

Ayers, M., Howell, P.L., and Burrows, L.L. (2010). Architecture of the type II secretion and type IV pilus machineries. *Future Microbiol* 5(8)**,** 1203-1218. doi: 10.2217/fmb.10.76.

Benz, I., and Schmidt, M.A. (1989). Cloning and expression of an adhesin (AIDA-I) involved in diffuse adherence of enteropathogenic *Escherichia coli*. *Infect Immun* 57(5)**,** 1506-1511.

Berger, C.N., Sodha, S.V., Shaw, R.K., Griffin, P.M., Pink, D., Hand, P., *et al.* (2010). Fresh fruit and vegetables as vehicles for the transmission of human pathogens. *Environ*

Microbiol 12(9), 2385-2397. doi: 10.1111/j.1462-2920.2010.02297.x.

Boisen, N., Melton-Celsa, A.R., Scheutz, F., O'Brien, A.D., and Nataro, J.P. (2015). Shiga toxin 2a and Enteroaggregative *Escherichia coli*--a deadly combination. *Gut Microbes* 6(4), 272-278. doi: 10.1080/19490976.2015.1054591.

Boisen, N., Struve, C., Scheutz, F., Krogfelt, K.A., and Nataro, J.P. (2008). New adhesin of enteroaggregative *Escherichia coli* related to the Afa/Dr/AAF family. *Infect Immun* 76(7), 3281-3292. doi: 10.1128/IAI.01646-07.

Brzuszkiewicz, E., Thurmer, A., Schuldes, J., Leimbach, A., Liesegang, H., Meyer, F.D., *et al.* (2011). Genome sequence analyses of two isolates from the recent *Escherichia coli* outbreak in Germany reveal the emergence of a new pathotype: Entero-Aggregative-Haemorrhagic *Escherichia coli* (EAHEC). *Arch Microbiol* 193(12), 883-891. doi: 10.1007/s00203-011-0725-6.

Busch, A., and Waksman, G. (2012). Chaperone-usher pathways: diversity and pilus assembly mechanism. *Philos Trans R Soc Lond B Biol Sci* 367(1592), 1112-1122. doi: 10.1098/rstb.2011.0206.

Chagnot, C., Zorgani, M.A., Astruc, T., and Desvaux, M. (2013). Proteinaceous determinants of surface colonization in bacteria: bacterial adhesion and biofilm formation from a protein secretion perspective. *Front Microbiol* 4, 303. doi: 10.3389/fmicb.2013.00303.

Choi, P.S., Dawson, A.J., and Bernstein, H.D. (2007). Characterization of a novel two-partner secretion system in *Escherichia coli* O157:H7. *J Bacteriol* 189(9), 3452-3461. doi: 10.1128/JB.01751-06.

Christie, P.J., and Vogel, J.P. (2000). Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol* 8(8), 354-360.

Christie, P.J., Whitaker, N., and Gonzalez-Rivera, C. (2014). Mechanism and structure of the bacterial type IV secretion systems. *Biochim Biophys Acta* 1843(8), 1578-1591. doi: 10.1016/j.bbamcr.2013.12.019.

Cookson, A.L., Cooley, W.A., and Woodward, M.J. (2002). The role of type 1 and curli fimbriae of Shiga toxin-producing *Escherichia coli* in adherence to abiotic surfaces. *Int J Med Microbiol* 292(3-4), 195-205. doi: 10.1078/1438-4221-00203.

Cornelis, G.R. (2006). The type III secretion injectisome. *Nat Rev Microbiol* 4(11), 811-825. doi: 10.1038/nrmicro1526.

Craig, L., Pique, M.E., and Tainer, J.A. (2004). Type IV pilus structure and bacterial pathogenicity. *Nat Rev Microbiol* 2(5), 363-378. doi: 10.1038/nrmicro885.

Croxen, M.A., Law, R.J., Scholz, R., Keeney, K.M., Wlodarska, M., and Finlay, B.B. (2013).

Recent advances in understanding enteric pathogenic Escherichia coli. *Clin Microbiol Rev* 26(4), 822-880. doi: 10.1128/CMR.00022-13.

Desvaux, M., Hebraud, M., Henderson, I.R., and Pallen, M.J. (2006). Type III secretion: what's in a name? *Trends Microbiol* 14(4), 157-160. doi: 10.1016/j.tim.2006.02.009.

Desvaux, M., Hebraud, M., Talon, R., and Henderson, I.R. (2009). Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. *Trends Microbiol* 17(4), 139-145. doi: 10.1016/j.tim.2009.01.004.

Desvaux, M., Parham, N.J., and Henderson, I.R. (2004). Type V protein secretion: simplicity gone awry? *Curr Issues Mol Biol* 6(2), 111-124.

Diepold, A., and Armitage, J.P. (2015). Type III secretion systems: the bacterial flagellum and the injectisome. *Philos Trans R Soc Lond B Biol Sci* 370(1679). doi: 10.1098/rstb.2015.0020.

DiLuzio, W.R., Turner, L., Mayer, M., Garstecki, P., Weibel, D.B., Berg, H.C., *et al.* (2005). *Escherichia coli* swim on the right-hand side. *Nature* 435(7046), 1271-1274. doi: 10.1038/nature03660.

Donnenberg, M.S., Tacket, C.O., James, S.P., Losonsky, G., Nataro, J.P., Wasserman, S.S., *et al.* (1993). Role of the eaeA gene in experimental enteropathogenic *Escherichia coli* infection. *J Clin Invest* 92(3), 1412-1417. doi: 10.1172/JCI116717.

Doughty, S., Sloan, J., Bennett-Wood, V., Robertson, M., Robins-Browne, R.M., and Hartland, E.L. (2002). Identification of a novel fimbrial gene cluster related to long polar fimbriae in locus of enterocyte effacement-negative strains of enterohemorrhagic *Escherichia coli*. *Infect Immun* 70(12), 6761-6769.

Dudley, E.G., Abe, C., Ghigo, J.M., Latour-Lambert, P., Hormazabal, J.C., and Nataro, J.P. (2006). An IncI1 plasmid contributes to the adherence of the atypical enteroaggregative Escherichia coli strain C1096 to cultured cells and abiotic surfaces. *Infect Immun* 74(4), 2102-2114. doi: 10.1128/IAI.74.4.2102-2114.2006.

Durand, E., Alphonse, S., Brochier-Armanet, C., Ball, G., Douzi, B., Filloux, A., *et al.* (2011). The assembly mode of the pseudopilus: a hallmark to distinguish a novel secretion system subtype. *J Biol Chem* 286(27), 24407-24416. doi: 10.1074/jbc.M111.234278.

Dziva, F., Mahajan, A., Cameron, P., Currie, C., McKendrick, I.J., Wallis, T.S., *et al.* (2007). EspP, a Type V-secreted serine protease of enterohaemorrhagic *Escherichia coli* O157:H7, influences intestinal colonization of calves and adherence to bovine primary intestinal epithelial cells. *FEMS Microbiol Lett* 271(2), 258-264. doi: 10.1111/j.1574-6968.2007.00724.x.

Easton, D.M., Allsopp, L.P., Phan, M.D., Moriel, D.G., Goh, G.K., Beatson, S.A., *et al.* (2014). The Intimin-Like Protein FdeC Is Regulated by H-NS and Temperature in Enterohemorrhagic *Escherichia coli*. *Appl Environ Microbiol* 80(23), 7337-7347. doi: 10.1128/AEM.02114-14.

Easton, D.M., Totsika, M., Allsopp, L.P., Phan, M.D., Idris, A., Wurpel, D.J., *et al.* (2011). Characterization of EhaJ, a New Autotransporter Protein from Enterohemorrhagic and Enteropathogenic *Escherichia coli*. *Front Microbiol* 2, 120. doi: 10.3389/fmicb.2011.00120.

El Tahir, Y., and Skurnik, M. (2001). YadA, the multifaceted *Yersinia* adhesin. *Int J Med Microbiol* 291(3), 209-218. doi: 10.1078/1438-4221-00119.

Enami, M., Nakasone, N., Honma, Y., Kakinohana, S., Kudaka, J., and Iwanaga, M. (1999). Expression of type I pili is abolished in verotoxin-producing *Escherichia coli* O157. *FEMS Microbiol Lett* 179(2), 467-472.

Farfan, M.J., Cantero, L., Vidal, R., Botkin, D.J., and Torres, A.G. (2011). Long polar fimbriae of enterohemorrhagic *Escherichia coli* O157:H7 bind to extracellular matrix proteins. *Infect Immun* 79(9), 3744-3750. doi: 10.1128/IAI.05317-11.

Farfan, M.J., Inman, K.G., and Nataro, J.P. (2008). The major pilin subunit of the AAF/II fimbriae from enteroaggregative *Escherichia coli* mediates binding to extracellular matrix proteins. *Infect Immun* 76(10), 4378-4384. doi: 10.1128/IAI.00439-08.

Farfan, M.J., and Torres, A.G. (2012). Molecular mechanisms that mediate colonization of Shiga toxin-producing *Escherichia coli* strains. *Infect Immun* 80(3), 903-913. doi: 10.1128/IAI.05907-11.

Finlay, B.B., and Falkow, S. (1997). Common themes in microbial pathogenicity revisited. *Microbiol Mol Biol Rev* 61(2), 136-169.

Fratamico, P.M., Yan, X., Caprioli, A., Esposito, G., Needleman, D.S., Pepe, T., *et al.* (2011). The complete DNA sequence and analysis of the virulence plasmid and of five additional plasmids carried by Shiga toxin-producing *Escherichia coli* O26:H11 strain H30. *Int J Med Microbiol* 301(3), 192-203. doi: 10.1016/j.ijmm.2010.09.002.

Fronzes, R., Christie, P.J., and Waksman, G. (2009). The structural biology of type IV secretion systems. *Nat Rev Microbiol* 7(10), 703-714. doi: 10.1038/nrmicro2218.

Galan, J.E., and Wolf-Watz, H. (2006). Protein delivery into eukaryotic cells by type III secretion machines. *Nature* 444(7119), 567-573. doi: 10.1038/nature05272.

Galfi, P., Neogrady, S., Semjen, G., Bardocz, S., and Pusztai, A. (1998). Attachment of different *Escherichia coli* strains to cultured rumen epithelial cells. *Vet Microbiol* 61(3), 191-197.

Gerlach, R.G., and Hensel, M. (2007). Protein secretion systems and adhesins: the molecular armory of Gram-negative pathogens. *Int J Med Microbiol* 297(6), 401-415. doi: 10.1016/j.ijmm.2007.03.017.

Ghigo, J.M. (2001). Natural conjugative plasmids induce bacterial biofilm development. *Nature* 412(6845), 442-445. doi: 10.1038/35086581.

Gonyar, L.A., and Kendall, M.M. (2014). Ethanolamine and choline promote expression of putative and characterized fimbriae in enterohemorrhagic *Escherichia coli* O157:H7. *Infect Immun* 82(1), 193-201. doi: 10.1128/IAI.00980-13.

Goyal, P., Krasteva, P.V., Van Gerven, N., Gubellini, F., Van den Broeck, I., Troupiotis-Tsailaki, A., *et al.* (2014). Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. *Nature* 516(7530), 250-253. doi: 10.1038/nature13768.

Haiko, J., and Westerlund-Wikstrom, B. (2013). The role of the bacterial flagellum in adhesion and virulence. *Biology (Basel)* 2(4), 1242-1267. doi: 10.3390/biology2041242.

Hammar, M., Bian, Z., and Normark, S. (1996). Nucleator-dependent intercellular assembly of adhesive curli organelles in *Escherichia coli*. *Proc Natl Acad Sci U S A* 93(13), 6562-6566.

Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., *et al.* (2001). Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res* 8(1), 11-22.

Henderson, I.R., Navarro-Garcia, F., Desvaux, M., Fernandez, R.C., and Ala'Aldeen, D. (2004). Type V protein secretion pathway: the autotransporter story. *Microbiol Mol Biol Rev* 68(4), 692-744. doi: 10.1128/MMBR.68.4.692-744.2004.

Herold, S., Paton, J.C., and Paton, A.W. (2009a). Sab, a novel autotransporter of locus of enterocyte effacement-negative shiga-toxigenic *Escherichia coli* O113:H21, contributes to adherence and biofilm formation. *Infect Immun* 77(8), 3234-3243. doi: 10.1128/IAI.00031-09.

Herold, S., Paton, J.C., Srimanote, P., and Paton, A.W. (2009b). Differential effects of short-chain fatty acids and iron on expression of iha in Shiga-toxigenic *Escherichia coli*. *Microbiology* 155(Pt 11), 3554-3563. doi: 10.1099/mic.0.029454-0.

Kelly, J.K., Pai, C.H., Jadusingh, I.H., Macinnis, M.L., Shaffer, E.A., and Hershfield, N.B. (1987). The histopathology of rectosigmoid biopsies from adults with bloody diarrhea due to verotoxin-producing *Escherichia coli*. *Am J Clin Pathol* 88(1), 78-82.

Klemm, P., Vejborg, R.M., and Sherlock, O. (2006). Self-associating autotransporters, SAATs: functional and structural similarities. *Int J Med Microbiol* 296(4-5), 187-195. doi:

10.1016/j.ijmm.2005.10.002.

Knowles, T.J., Scott-Tucker, A., Overduin, M., and Henderson, I.R. (2009). Membrane protein architects: the role of the BAM complex in outer membrane protein assembly. *Nat Rev Microbiol* 7(3)**,** 206-214. doi: 10.1038/nrmicro2069.

Korea, C.G., Badouraly, R., Prevost, M.C., Ghigo, J.M., and Beloin, C. (2010). *Escherichia coli* K-12 possesses multiple cryptic but functional chaperone-usher fimbriae with distinct surface specificities. *Environ Microbiol* 12(7)**,** 1957-1977. doi: 10.1111/j.1462-2920.2010.02202.x.

Lawley, T.D., Klimke, W.A., Gubbins, M.J., and Frost, L.S. (2003). F factor conjugation is a true type IV secretion system. *FEMS Microbiol Lett* 224(1)**,** 1-15.

Leo, J.C., Oberhettinger, P., Schutz, M., and Linke, D. (2015). The inverse autotransporter family: intimin, invasin and related proteins. *Int J Med Microbiol* 305(2)**,** 276-282. doi: 10.1016/j.ijmm.2014.12.011.

Lewis, S.B., Cook, V., Tighe, R., and Schuller, S. (2015). Enterohemorrhagic Escherichia coli colonization of human colonic epithelium in vitro and ex vivo. *Infect Immun* 83(3)**,** 942-949. doi: 10.1128/IAI.02928-14.

Li, B., Koch, W.H., and Cebula, T.A. (1997). Detection and characterization of the fimA gene of *Escherichia coli* O157:H7. *Mol Cell Probes* 11(6)**,** 397-406. doi: 10.1006/mcpr.1997.0132.

Lim, J.Y., Yoon, J., and Hovde, C.J. (2010). A brief overview of *Escherichia coli* O157:H7 and its plasmid O157. *J Microbiol Biotechnol* 20(1)**,** 5-14.

Linke, D., Riess, T., Autenrieth, I.B., Lupas, A., and Kempf, V.A. (2006). Trimeric autotransporter adhesins: variable structure, common function. *Trends Microbiol* 14(6)**,** 264-270. doi: 10.1016/j.tim.2006.04.005.

Liu, H., Magoun, L., and Leong, J.M. (1999a). beta1-chain integrins are not essential for intimin-mediated host cell attachment and enteropathogenic *Escherichia coli*-induced actin condensation. *Infect Immun* 67(4)**,** 2045-2049.

Liu, H., Magoun, L., Luperchio, S., Schauer, D.B., and Leong, J.M. (1999b). The Tir-binding region of enterohaemorrhagic *Escherichia coli* intimin is sufficient to trigger actin condensation after bacterial-induced host cell signalling. *Mol Microbiol* 34(1)**,** 67-81.

Lloyd, S.J., Ritchie, J.M., Rojas-Lopez, M., Blumentritt, C.A., Popov, V.L., Greenwich, J.L., *et al.* (2012). A double, long polar fimbria mutant of *Escherichia coli* O157:H7 expresses Curli and exhibits reduced in vivo colonization. *Infect Immun* 80(3)**,** 914-920. doi: 10.1128/IAI.05945-11.

Low, A.S., Dziva, F., Torres, A.G., Martinez, J.L., Rosser, T., Naylor, S., *et al.* (2006a). Cloning, expression, and characterization of fimbrial operon F9 from enterohemorrhagic *Escherichia coli* O157:H7. *Infect Immun* 74(4)**,** 2233-2244. doi: 10.1128/IAI.74.4.2233-2244.2006.

Low, A.S., Holden, N., Rosser, T., Roe, A.J., Constantinidou, C., Hobman, J.L., *et al.* (2006b). Analysis of fimbrial gene clusters and their expression in enterohaemorrhagic *Escherichia coli* O157:H7. *Environ Microbiol* 8(6)**,** 1033-1047. doi: 10.1111/j.1462-2920.2006.00995.x.

Lu, Y., Iyoda, S., Satou, H., Satou, H., Itoh, K., Saitoh, T., *et al.* (2006). A new immunoglobulin-binding protein, EibG, is responsible for the chain-like adhesion phenotype of locus of enterocyte effacement-negative, shiga toxin-producing *Escherichia coli*. *Infect Immun* 74(10)**,** 5747-5755. doi: 10.1128/IAI.00724-06.

Macnab, R.M. (2004). Type III flagellar protein export and flagellar assembly. *Biochim Biophys Acta* 1694(1-3)**,** 207-217. doi: 10.1016/j.bbamcr.2004.04.005.

Mahajan, A., Currie, C.G., Mackie, S., Tree, J., McAteer, S., McKendrick, I., *et al.* (2009). An investigation of the expression and adhesin function of H7 flagella in the interaction of *Escherichia coli* O157 : H7 with bovine intestinal epithelium. *Cell Microbiol* 11(1)**,** 121-137. doi: 10.1111/j.1462-5822.2008.01244.x.

Mallick, E.M., Brady, M.J., Luperchio, S.A., Vanguri, V.K., Magoun, L., Liu, H., *et al.* (2012). Allele- and tir-independent functions of intimin in diverse animal infection models. *Front Microbiol* 3**,** 11. doi: 10.3389/fmicb.2012.00011.

May, T., Ito, A., and Okabe, S. (2010). Characterization and global gene expression of F-phenocopies during *Escherichia coli* biofilm formation. *Mol Genet Genomics* 284(5)**,** 333-342. doi: 10.1007/s00438-010-0571-2.

May, T., and Okabe, S. (2008). *Escherichia coli* harboring a natural IncF conjugative F plasmid develops complex mature biofilms by stimulating synthesis of colanic acid and Curli. *J Bacteriol* 190(22)**,** 7479-7490. doi: 10.1128/JB.00823-08.

McDaniel, T.K., Jarvis, K.G., Donnenberg, M.S., and Kaper, J.B. (1995). A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc Natl Acad Sci U S A* 92(5)**,** 1664-1668.

McNeilly, T.N., Naylor, S.W., Mahajan, A., Mitchell, M.C., McAteer, S., Deane, D., *et al.* (2008). *Escherichia coli* O157:H7 colonization in cattle following systemic and mucosal immunization with purified H7 flagellin. *Infect Immun* 76(6)**,** 2594-2602. doi: 10.1128/IAI.01452-07.

Mellmann, A., Harmsen, D., Cummings, C.A., Zentz, E.B., Leopold, S.R., Rico, A., *et al.* (2011). Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS One* 6(7), e22751. doi: 10.1371/journal.pone.0022751.

Merkel, V., Ohder, B., Bielaszewska, M., Zhang, W., Fruth, A., Menge, C., *et al.* (2010). Distribution and phylogeny of immunoglobulin-binding protein G in Shiga toxin-producing *Escherichia coli* and its association with adherence phenotypes. *Infect Immun* 78(8), 3625-3636. doi: 10.1128/IAI.00006-10.

Musken, A., Bielaszewska, M., Greune, L., Schweppe, C.H., Muthing, J., Schmidt, H., *et al.* (2008). Anaerobic conditions promote expression of Sfp fimbriae and adherence of sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157:NM to human intestinal epithelial cells. *Appl Environ Microbiol* 74(4), 1087-1093. doi: 10.1128/AEM.02496-07.

Nataro, J.P., and Kaper, J.B. (1998). Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev* 11(1), 142-201.

Nesta, B., Spraggon, G., Alteri, C., Moriel, D.G., Rosini, R., Veggi, D., *et al.* (2012). FdeC, a novel broadly conserved *Escherichia coli* adhesin eliciting protection against urinary tract infections. *MBio* 3(2). doi: 10.1128/mBio.00010-12.

Oberhettinger, P., Leo, J.C., Linke, D., Autenrieth, I.B., and Schutz, M.S. (2015). The inverse autotransporter intimin exports its passenger domain via a hairpin intermediate. *J Biol Chem* 290(3), 1837-1849. doi: 10.1074/jbc.M114.604769.

Oberhettinger, P., Schutz, M., Leo, J.C., Heinz, N., Berger, J., Autenrieth, I.B., *et al.* (2012). Intimin and invasin export their C-terminus to the bacterial cell surface using an inverse mechanism compared to classical autotransport. *PLoS One* 7(10), e47069. doi: 10.1371/journal.pone.0047069.

Ostroff, S.M., Tarr, P.I., Neill, M.A., Lewis, J.H., Hargrett-Bean, N., and Kobayashi, J.M. (1989). Toxin genotypes and plasmid profiles as determinants of systemic sequelae in *Escherichia coli* O157:H7 infections. *J Infect Dis* 160(6), 994-998.

Oswald, E., Schmidt, H., Morabito, S., Karch, H., Marches, O., and Caprioli, A. (2000). Typing of intimin genes in human and animal enterohemorrhagic and enteropathogenic *Escherichia coli:* characterization of a new intimin variant. *Infect Immun* 68(1), 64-71.

Paros, M., Tarr, P.I., Kim, H., Besser, T.E., and Hancock, D.D. (1993). A comparison of human and bovine *Escherichia coli* O157:H7 isolates by toxin genotype, plasmid profile, and bacteriophage lambda-restriction fragment length polymorphism profile. *J Infect Dis*

168(5), 1300-1303.

Paton, A.W., Srimanote, P., Woodrow, M.C., and Paton, J.C. (2001). Characterization of Saa, a novel autoagglutinating adhesin produced by *locus* of enterocyte effacement-negative Shiga-toxigenic *Escherichia coli* strains that are virulent for humans. *Infect Immun* 69(11), 6999-7009. doi: 10.1128/IAI.69.11.6999-7009.2001.

Puttamreddy, S., Cornick, N.A., and Minion, F.C. (2010). Genome-wide transposon mutagenesis reveals a role for pO157 genes in biofilm development in *Escherichia coli* O157:H7 EDL933. *Infect Immun* 78(6), 2377-2384. doi: 10.1128/IAI.00156-10.

Rendon, M.A., Saldana, Z., Erdem, A.L., Monteiro-Neto, V., Vazquez, A., Kaper, J.B., *et al.* (2007). Commensal and pathogenic Escherichia coli use a common pilus adherence factor for epithelial cell colonization. *Proc Natl Acad Sci U S A* 104(25), 10637-10642. doi: 0704104104 [pii]10.1073/pnas.0704104104.

Roe, A.J., Currie, C., Smith, D.G., and Gally, D.L. (2001). Analysis of type 1 fimbriae expression in verotoxigenic *Escherichia coli*: a comparison between serotypes O157 and O26. *Microbiology* 147(Pt 1), 145-152. doi: 10.1099/00221287-147-1-145.

Ross, B.N., Rojas-Lopez, M., Cieza, R.J., McWilliams, B.D., and Torres, A.G. (2015). The Role of Long Polar Fimbriae in *Escherichia coli* O104:H4 Adhesion and Colonization. *PLoS One* 10(10), e0141845. doi: 10.1371/journal.pone.0141845.

Rossez, Y., Holmes, A., Lodberg-Pedersen, H., Birse, L., Marshall, J., Willats, W.G., *et al.* (2014). *Escherichia coli* common pilus (ECP) targets arabinosyl residues in plant cell walls to mediate adhesion to fresh produce plants. *J Biol Chem* 289(49), 34349-34365. doi: 10.1074/jbc.M114.587717.

Ryu, J.H., and Beuchat, L.R. (2005). Biofilm formation by *Escherichia coli* O157:H7 on stainless steel: effect of exopolysaccharide and Curli production on its resistance to chlorine. *Appl Environ Microbiol* 71(1), 247-254. doi: 10.1128/AEM.71.1.247-254.2005.

Saldana, Z., Erdem, A.L., Schuller, S., Okeke, I.N., Lucas, M., Sivananthan, A., *et al.* (2009). The *Escherichia coli* common pilus and the bundle-forming pilus act in concert during the formation of localized adherence by enteropathogenic *E. coli*. *J Bacteriol* 191(11), 3451-3461. doi: 10.1128/JB.01539-08.

Samadder, P., Xicohtencatl-Cortes, J., Saldana, Z., Jordan, D., Tarr, P.I., Kaper, J.B., *et al.* (2009). The *Escherichia coli* ycbQRST operon encodes fimbriae with laminin-binding and epithelial cell adherence properties in Shiga-toxigenic *E. coli* O157:H7. *Environ Microbiol* 11(7), 1815-1826. doi: 10.1111/j.1462-2920.2009.01906.x.

Schmidt, M.A. (2010). LEEways: tales of EPEC, ATEC and EHEC. *Cell Microbiol* 12(11), 1544-1552. doi: 10.1111/j.1462-5822.2010.01518.x.

Schroeder, G.N., and Hilbi, H. (2008). Molecular pathogenesis of *Shigella* spp.: controlling host cell signaling, invasion, and death by type III secretion. *Clin Microbiol Rev* 21(1), 134-156. doi: 10.1128/CMR.00032-07.

Shaw, R.K., Berger, C.N., Feys, B., Knutton, S., Pallen, M.J., and Frankel, G. (2008). Enterohemorrhagic *Escherichia coli* exploits EspA filaments for attachment to salad leaves. *Appl Environ Microbiol* 74(9), 2908-2914. doi: 10.1128/AEM.02704-07.

Sherlock, O., Schembri, M.A., Reisner, A., and Klemm, P. (2004). Novel roles for the AIDA adhesin from diarrheagenic *Escherichia coli*: cell aggregation and biofilm formation. *J Bacteriol* 186(23), 8058-8065. doi: 10.1128/JB.186.23.8058-8065.2004.

Szczesny, P., and Lupas, A. (2008). Domain annotation of trimeric autotransporter adhesins--daTAA. *Bioinformatics* 24(10), 1251-1256. doi: 10.1093/bioinformatics/btn118.

Tarr, P.I., Bilge, S.S., Vary, J.C., Jr., Jelacic, S., Habeeb, R.L., Ward, T.R., *et al.* (2000). Iha: a novel *Escherichia coli* O157:H7 adherence-conferring molecule encoded on a recently acquired chromosomal island of conserved structure. *Infect Immun* 68(3), 1400-1407.

Thanassi, D.G., Bliska, J.B., and Christie, P.J. (2012). Surface organelles assembled by secretion systems of Gram-negative bacteria: diversity in structure and function. *FEMS Microbiol Rev* 36(6), 1046-1082. doi: 10.1111/j.1574-6976.2012.00342.x.

Tietze, E., Dabrowski, P.W., Prager, R., Radonic, A., Fruth, A., Aurass, P., *et al.* (2015). Comparative genomic analysis of two novel sporadic Shiga toxin-producing *Escherichia coli* O104:H4 strains isolated 2011 in Germany. *PLoS One* 10(4), e0122074. doi: 10.1371/journal.pone.0122074.

Tjalsma, H., Bolhuis, A., Jongbloed, J.D., Bron, S., and van Dijl, J.M. (2000). Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiol Mol Biol Rev* 64(3), 515-547.

Torres, A.G., Jeter, C., Langley, W., and Matthysse, A.G. (2005). Differential binding of *Escherichia coli* O157:H7 to alfalfa, human epithelial cells, and plastic is mediated by a variety of surface structures. *Appl Environ Microbiol* 71(12), 8008-8015. doi: 10.1128/AEM.71.12.8008-8015.2005.

Torres, A.G., and Kaper, J.B. (2003). Multiple elements controlling adherence of enterohemorrhagic *Escherichia coli* O157:H7 to HeLa cells. *Infect Immun* 71(9), 4985-4995.

Torres, A.G., Slater, T.M., Patel, S.D., Popov, V.L., and Arenas-Hernandez, M.M. (2008).

Contribution of the Ler- and H-NS-regulated long polar fimbriae of *Escherichia coli* O157:H7 during binding to tissue-cultured cells. *Infect Immun* 76(11), 5062-5071. doi: 10.1128/IAI.00654-08.

Totsika, M., Wells, T.J., Beloin, C., Valle, J., Allsopp, L.P., King, N.P., *et al.* (2012). Molecular characterization of the EhaG and UpaG trimeric autotransporter proteins from pathogenic *Escherichia coli*. *Appl Environ Microbiol* 78(7), 2179-2189. doi: 10.1128/AEM.06680-11.

Turner, L., Zhang, R., Darnton, N.C., and Berg, H.C. (2010). Visualization of Flagella during bacterial Swarming. *J Bacteriol* 192(13), 3259-3267. doi: 10.1128/JB.00083-10.

Uhlich, G.A., Keen, J.E., and Elder, R.O. (2002). Variations in the csgD promoter of *Escherichia coli* O157:H7 associated with increased virulence in mice and increased invasion of HEp-2 cells. *Infect Immun* 70(1), 395-399.

van Diemen, P.M., Dziva, F., Stevens, M.P., and Wallis, T.S. (2005). Identification of enterohemorrhagic *Escherichia coli* O26:H- genes required for intestinal colonization in calves. *Infect Immun* 73(3), 1735-1743. doi: 10.1128/IAI.73.3.1735-1743.2005.

Voulhoux, R., Ball, G., Ize, B., Vasil, M.L., Lazdunski, A., Wu, L.F., *et al.* (2001). Involvement of the twin-arginine translocation system in protein secretion via the type II pathway. *EMBO J* 20(23), 6735-6741. doi: 10.1093/emboj/20.23.6735.

Wang, P., Xiong, Y., Lan, R., Ye, C., Wang, H., Ren, J., *et al.* (2011). pO157_Sal, a novel conjugative plasmid detected in outbreak isolates of *Escherichia coli* O157:H7. *J Clin Microbiol* 49(4), 1594-1597. doi: 10.1128/JCM.02530-10.

Wells, T.J., McNeilly, T.N., Totsika, M., Mahajan, A., Gally, D.L., and Schembri, M.A. (2009). The *Escherichia coli* O157:H7 EhaB autotransporter protein binds to laminin and collagen I and induces a serum IgA response in O157:H7 challenged cattle. *Environ Microbiol* 11(7), 1803-1814. doi: 10.1111/j.1462-2920.2009.01905.x.

Wells, T.J., Sherlock, O., Rivas, L., Mahajan, A., Beatson, S.A., Torpdahl, M., *et al.* (2008). EhaA is a novel autotransporter protein of enterohemorrhagic *Escherichia coli* O157:H7 that contributes to adhesion and biofilm formation. *Environ Microbiol* 10(3), 589-604. doi: 10.1111/j.1462-2920.2007.01479.x.

Wong, A.R., Pearson, J.S., Bright, M.D., Munera, D., Robinson, K.S., Lee, S.F., *et al.* (2011). Enteropathogenic and enterohaemorrhagic *Escherichia coli*: even more subversive elements. *Mol Microbiol* 80(6), 1420-1438. doi: 10.1111/j.1365-2958.2011.07661.x.

Xicohtencatl-Cortes, J., Monteiro-Neto, V., Saldana, Z., Ledesma, M.A., Puente, J.L., and Giron, J.A. (2009). The type 4 pili of enterohemorrhagic Escherichia coli O157:H7 are

multipurpose structures with pathogenic attributes. *J Bacteriol* 191(1)**,** 411-421. doi: JB.01306-08 [pii]10.1128/JB.01306-08.

Xicohtencatl-Cortes, J., Saldana, Z., Deng, W., Castaneda, E., Freer, E., Tarr, P.I., *et al.* (2010). Bacterial macroscopic rope-like fibers with cytopathic and adhesive properties. *J Biol Chem* 285(42)**,** 32336-32342. doi: 10.1074/jbc.M110.162248.

Zav'yalov, V., Zavialov, A., Zav'yalova, G., and Korpela, T. (2010). Adhesive organelles of Gram-negative pathogens assembled with the classical chaperone/usher machinery: structure and function from a clinical standpoint. *FEMS Microbiol Rev* 34(3)**,** 317-378. doi: 10.1111/j.1574-6976.2009.00201.x.

Conclusions

Bouzari, S., Dashti, A., Jafari, A., and Oloomi, M. (2010). Immune response against adhesins of enteroaggregative Escherichia coli immunized by three different vaccination strategies (DNA/DNA, Protein/Protein, and DNA/Protein) in mice. Comp Immunol Microbiol Infect Dis 33, 215-225. 10.1016/j.cimid.2008.10.002

Cdc. Available: http://www.cdc.gov/onehealth/basics/index.html [Accessed].

Garcia-Angulo, V.A., Kalita, A., Kalita, M., Lozano, L., and Torres, A.G. (2014). Comparative genomics and immunoinformatics approach for the identification of vaccine candidates for enterohemorrhagic Escherichia coli O157:H7. *Infect Immun* 82(5)**,** 2016-2026. doi: 10.1128/IAI.01437-13.

Etcheverría, A.I., Lucchesi, P.M.A., Krüger, A., Bentancor, A.B., and Padola, N.L. (2016). "Escherichia coli in Animals," in Escherichia coli in the Americas, ed. A.G. Torres. (Cham: Springer International Publishing), 149-172.

Kulp, A., and Kuehn, M.J. (2010). Biological functions and biogenesis of secreted bacterial outer membrane vesicles. *Annu Rev Microbiol* 64**,** 163-184. 10.1146/annurev.micro.091208.073413

Rappuoli, R., Bottomley, M.J., D'Oro, U., Finco, O., and De Gregorio, E. (2016). Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design. J Exp Med 213(4), 469-481. doi: 10.1084/jem.20151960.

Schwechheimer, C., and Kuehn, M.J. (2015). Outer-membrane vesicles from Gram-negative bacteria: biogenesis and functions. Nat Rev Microbiol 13, 605-619.

10.1038/nrmicro3525

Torres, A.G. (2017). Escherichia coli diseases in Latin America-a 'One Health' multidisciplinary approach. Pathog Dis 75. 10.1093/femspd/ftx012

Tapia, D., Ross, B.N., Kalita, A., Kalita, M., Hatcher, C.L., Muruato, L.A., *et al.* (2016). From In silico Protein Epitope Density Prediction to Testing Escherichia coli O157:H7 Vaccine Candidates in a Murine Model of Colonization. *Front Cell Infect Microbiol* 6**,** 94. doi: 10.3389/fcimb.2016.00094.

Annexes

Annexe 1
Intestinal pathogenic *Escherichia coli*: insights for vaccines development

Maricarmen ROJAS-LOPEZ [1,2], Ricardo MONTERIO[2], Mariagrazia PIZZA [1], Mickaël DESVAUX [2,*] and Roberto ROSINI [1,*]

[1] GlaxoSmithKline Vaccines, Via Fiorentina 1, 53100, Siena, Italy.

[2] Université Clermont Auvergne, INRA, UMR454 MEDiS, 63000 Clermont-Ferrand, France.

*Corresponding author: Roberto ROSINI,. GlaxoSmithKline Vaccines srl., Via Fiorentina 1, 53100, Siena, Italy Tel.: +39 05 77 25 38 88, Fax: +39 05 77 25 38 88, e-Mail : roberto.x.rosini@gsk.com. Mickaël DESVAUX, Institut National de la Recherche Agronomique Auvergne-Rhône-Alpes, Centre de Recherche de Clermont-Ferrand, UMR454 MEDiS (Microbiologie, Environnement Digestif, Santé), Site de Theix, 63122 Saint-Genès Champanelle, France. Tel.: +33 (0)4 73 62 47 23, Fax: +33 (0)4 73 62 45 81, e-mail: mickael.desvaux@inra.fr

ABSTRACT

Diarrheal diseases are one of the major causes of mortality among under-five year old children and the intestinal pathogenic Escherichia coli (InPEC) are one of the large causative groups oh these diarrheal diseases worldwide. In this matter, InPECs contribute significantly to the burden of intestinal diseases, which are a critical issue in Low-Mid income countries (in Asia, Africa and Latin America). Intestinal pathotypes such as Enteropathogenic *E. coli* (EPEC) and enterotoxigenic *E. coli* (ETEC) are mainly endemic in developing countries, while ETEC strains are the major cause of diarrhoea in travelers to these countries. On the other hand, enterohemorrhagic *E. coli* (EHEC) are the cause of large outbreaks around the world mainly affecting developed countries and responsible not only of diarrheal disease but also of severe clinical complications like hemorrhagic colitis and hemolytic uremic syndrome (HUS). Overall, the emergence antibiotic resistance strains, the annual cost increase in the health care system, the high incidence in traveler diarrhea and increased number of HUS episodes, has arised the need of effective preventive treatments. Although the use of antibiotic is still an important factor

to treat such infections, non-antibiotic strategies are either a crucial option to limit the increase in antibiotic resistant strains or absolutely necessary for diseases, such as those caused by EHEC infections, for which antibiotic therapies are not recommended. Among non-antibiotic therapies, vaccine development is a strategy of choice but, to date, there is no licensed vaccine effective against InPEC infections. During several years, there has been an important effort to identify efficacious vaccine candidates able to assuage the diarrheal disease burden. The aim of this review is to summarize recent milestones and insights for vaccine development against InPECs.

KEY WORDS: Vaccines; InPEC: intestinal pathogenic *E. coli*; EHEC Enterohemorrhagic *E. coli*; Enterotoxigenic *E. coli*; Enteropathogenic *E. coli*.

INTRODUCTION

Escherichia coli is a Gram-negative commonly found as a commensal in the human microbiota. However, the plasticity of its genome has led to the evolution of this organism into pathogenic strains able to cause disease and syndromes of health importance in humans and animals. Pathogenic *E. coli* are mainly gathered into two groups depending of the disease localization: extraintestinal pathogenic *E. coli* (ExPEC) and intestinal pathogenic *E. coli* (InPEC). While ExPEC strains are associated mainly to neonatal meningitis (NMEC) and urinary tract infections (UPEC) in adults, InPEC strains, related to diarrheal disease, are subdivided in at least 6 well-known pathotypes: enteropathogenic *E. coli* (EPEC), enterohemorrhagic *E. coli* (EHEC), enteroinvasive *E. coli* (EIEC), enteroaggregative *E. coli* (EAEC), diffusely adherent *E. coli* (DAEC), and enterotoxigenic *E. coli* (ETEC). These *E. coli* enteropathotypes are classified according to their virulence factors, mechanisms of infection, interaction with the enterocyte, tissue tropism, symptoms and syndromes (Kaper *et al.*, 2004; Croxen and Finlay, 2010). Most recently, adherent-invasive *E. coli* (AIEC) have also been described as a disease-associated *E. coli* related to Crohn's disease but they cannot be *stricto sensu* considered as a diarrheagenic *E. coli* and thus classified as a new InPEC pathotype.

Nowadays, the burden of diarrheal diseases is one of the major problems in public health around the world caused by several etiological agents without the availability of preventive strategies such as vaccines to block infections. Besides InPEC, these diarrheagenic agents include *Cryptosporidium, Campylobacter, Shigella* and *Salmonella*, responsible for the vast majority of diarrheal diseases worldwide (MacLennan and Saul, 2014). Furthermore, in order to standardize the real estimation of the burden of diarrheal disease, the etiology and the real

number of the most affected populations, mathematical approaches were developed. Specifically, to calculate numbers in the burden of diarrheal disease in children under 5-years old (U5), focused not only on diarrheal diseases but also on other diseases as pneumonia the Global Burden of Disease (GBD by Health Metrics and Evaluation, 2010) and Child Health Epidemiology Reference Group (CHERG by World Health Organization) (Kovacs *et al.*, 2015) indexes were implemented. The main goal of these approaches is the estimation of the cause-specific disease and morbidity-mortality, but still the output brings to different numbers due the diverse inclusion data included into these methods. In addition, the Global Enteric Multicenter Study (GEMS), a case-control study still ongoing in Africa and Asia, is aimed to identify the etiology and population based burden of pediatric diarrheal disease (Kotloff *et al.*, 2013).

Overall, the use of these approaches will allowed to refine the regulations, therapies, control strategies and to estimate the real numbers of the infectious diseases depending on the specific target (Pires *et al.*, 2015).

In this matter, InPECs contribute significantly to the burden of diarrheal diseases, which represent a critical issue in Low-Mid income countries (in Asia, Africa and Latin America). For instance, EPEC and ETEC are mainly endemic in developing countries and the last in travelers to these countries (Kaper *et al.*, 2004; Torres, 2017a), whereas EHEC are responsible for large outbreaks around the world. This pathotype, mainly affecting developed countries, causes not only diarrheal disease but is responsible for clinical complications like hemorrhagic colitis and hemolytic uremic syndrome (HUS), a rising issue in Latin-American countries like Argentina (Kaper *et al.*, 2004; Pianciola *et al.*, 2016; Torres, 2017a). Similarly, EAEC has also been involved occasionally in diarrheal diseases in developing and industrialized countries (Foster *et al.*, 2015) and in 2011 an EHEC-EAEC hybrid has emerged causing a large outbreak in Europe, with 3816 reported cases, leading to 845 HUS cases and 54 deaths (Brzuszkiewicz *et al.*, 2011; Frank *et al.*, 2011).

Overall, *(i)* the increase in the burden of the disease as a high incidence in traveler diarrhea (ETEC, EAEC), *(ii)* endemic ETEC and EPEC cases in developing countries, *(iii)* infections with EHEC and increased number of HUS episodes, *(iv)* the annual cost for the healthcare system, *(v)* the emerging of antibiotic resistance strains, arises the need of an effective preventive treatment to assuage the burden of the diarrheal disease. Although the use of antibiotic is still key to treat such infections, non-antibiotic strategies are either a crucial option to limit the increase in antibiotic resistant strains, (Torres, 2017a) or the unique option for diseases where antibiotic therapies are not recommended, e.g. EHEC infection (Goldwater and Bettelheim, 2012; Rivas *et al.*, 2016) . Among non-antibiotic therapies, vaccine development

is a strategy of choice but, to date, there is no universal or specific licensed vaccine against InPEC.

For years the vaccine development has included different platforms and approaches as (*i)* pathogens attenuated by different environmental conditions exposure (heat or oxygen) or by multiple passages in media culture (*in vitro*) a method considered as the most ancient and empirical form of vaccine production ; (*ii)* detoxified toxins forms, like the detoxified version of the diphtheria and tetanus toxin; (*iii)* the use of protein based vaccines, as the hemagglutinin from influenza virus or the vaccine for *Bordetella pertussis*; (*iv)* genetic engineered vaccines, that have been the most exploited alternative to vaccine development, in which antigens can be produced in different vectors that would reduce the toxicity or collateral immunoreactions (Mora *et al.*, 2003; Plotkin, 2014).

However, there are several infectious diseases for which these traditional approaches have failed and for which vaccines have not yet been developed. With the advent of whole-genome sequencing and advances in bioinformatics, the vaccinology field has changed. Approaches like reverse vaccinology, based on the scanning of the annotate complete pathogen genome by bioinformatic prediction of the most likely vaccines candidates, have let the identification of promising antigens and development of a safe broad protective vaccine against the *Neisseria meningitides* serogroup B pathogen (Pizza *et al.*, 2000; Giuliani *et al.*, 2006; Feavers and Maiden, 2017). For other pathogens such as *Streptococcus ssp*., and for extra intestinal pathogenic *E. coli* (ExPEC) a number of promising antigens have also been identified (De Gregorio and Rappuoli, 2012; Sjoling *et al.*, 2015). In particular, for ExPEC the genome sequence analysis of a neonatal meningitis isolate (NMEC) let the identification of 230 potential antigens. The most protective antigens uncovered from that analysis, were a broadly conserved adhesin (FdeC,) and a conserved secreted zinc metallopeptidase (SslE,), which conferred cross protection in three different murine models including intestinal, ascending urinary tract infection and sepsis model (Moriel *et al.*, 2010; Nesta *et al.*, 2012; Nesta *et al.*, 2014).

During several years, there has been significant efforts to identify effective vaccines t against InPEC by different research groups worldwide. Up to date the major pathotypes considered for the vaccine development are EHEC and ETEC for their main impact in the burden in the public health (Table 1); even though EPEC is a major issue in developing countries there are not considerable advances on it. The aim of this review is to summarize the advances made in the development of vaccines against these different InPEC pathotypes and the perspectives they offer (Figure 1).

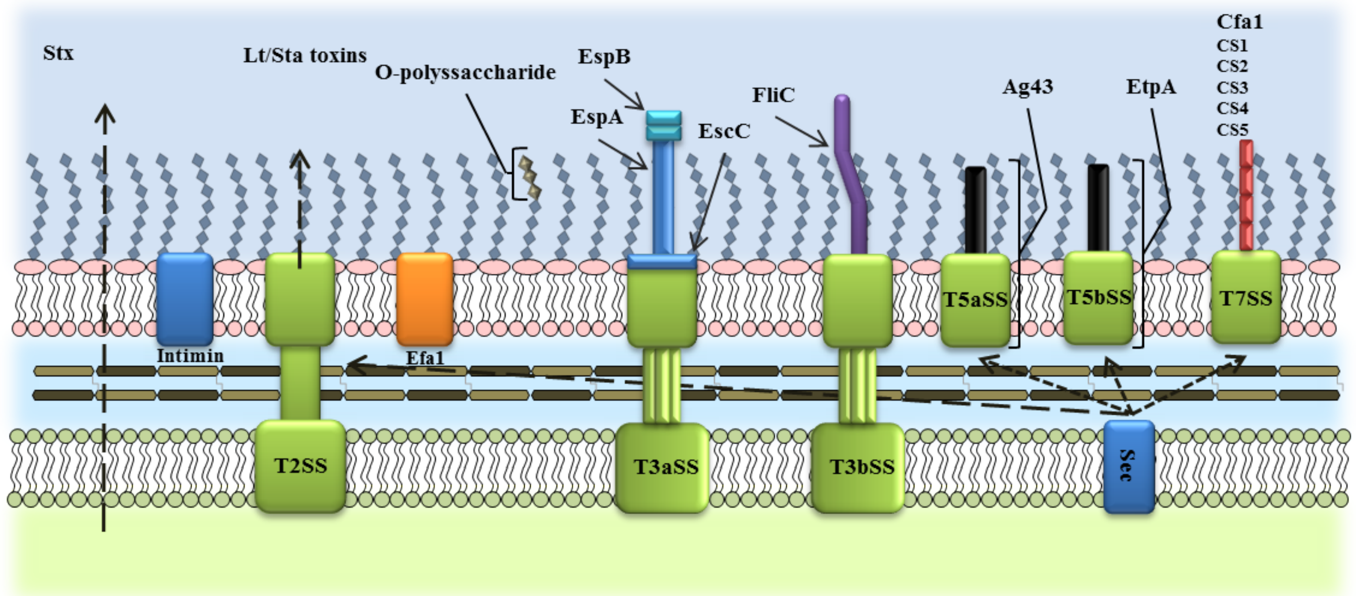


Figure 1. Schematic respresentation of the viruelence factors used as vaccine candidates. Several of the InPECs virulence factors have been employed as vaccine candidates, including outer membrane proteins, toxins, O-polyssaccharides, exported and secreted proteins. T2SS: Type two secretion system; T3aSS: Type three A secretion system; T3bSS: Type three B secretion system; T5aSS: Type five A secretion system; T5aSS: Type five B secretion system; T7SS: Type seven secretion system.

VACCINE DEVELOPMENT AGAINST EHEC

As for each of the characterized InPEC, EHEC are primarily defined on an array of evidences based on clinical manifestations, i.e. clinical symptoms, histological and molecular features (Nataro and Kaper, 1998; Kaper *et al.*, 2004). One of its primary feature is the presence of Shiga-toxin (Stx) genes; EHEC thus belong to the genotypic group of Shiga-toxin encoding *E. coli* (STEC). While there are over 380 distinct serotypes of STEC (Karmali *et al.*, 2003; Karmali *et al.*, 2010), only EHEC of serotypes O157:H7, as well as the big six non-O157 serotypes O26:H11, O45:H2, O103:H2, O111:H8, O121:H19 and O145:H28 are the most frequently associated with human disease and large outbreaks around the world (Karmali *et al.*, 2003; Karmali *et al.*, 2010). Based on their incidence, involvement in outbreaks and association with severe disease, STEC can be classified into 4 classes, namely *(i)* class A corresponding to the O157 serogroup with a high virulence, *(ii)* class B to the big six non-O157 serotypes mentioned above with a moderate virulence, *(iii)* class C to serotypes mainly involved in some sporadic cases with a moderate virulence, and *(iv)* class D to serotypes never reported in human infection

and thus avirulent (Karmali *et al.*, 2003). In this regard, EHEC clearly defined a subset of InPEC clinically isolated but STEC do not systematically and cannot definitely refer to a specific intestinal pathogenic *E. coli* pathotype (Nataro and Kaper, 1998). Currently, the question if a STEC strain isolated from the environment or food for instance is truly an EHEC, if itis virulent (and at which level) and a risk for human health, cannot be simply answered (Messens *et al.*, 2015). In the United State by 2014 the number of cases per 100,000 population were 690 for EHEC O157 and 445 for EHEC non-O157 while percentage of infections associated with outbreaks accounted for 16% for EHEC O157 and 7% for EHEC non-O157 (Stacy M. Crim and Matthew Cartter, 2015). Regarding EHEC non-O157:H7, their incidence is on the increase and became important human pathogens as important as EHEC O157:H7 (Croxen *et al.*, 2013). More recently the infection with EHEC strains is rising in Latin American countries, becoming an endemic phenomenon, as Argentina (Mejias *et al.*, 2016).

As asymptomatical carrier, cattle are the natural reservoir of EHEC and as an anthropozoonose, infection can occur from direct (meat, milk) or indirect (vegetables, water) contamination of food products from animal feces (Rivas *et al.*, 2016). The adhesion and persistence of the bacteria is an important key feature in the bacterial pathogenesis. In fact, EHEC O157:H7 colonize mainly human colon and carry a pathogenic island (PAI) known as Locus of Enterocyte Effacement (LEE) (like EPEC) that encodes the type three secretion system (T3SS) and it would give the capability to produce the A/E lesion in humans, but this phenomenon is mainly observed in *in vitro* assays than *in vivo*. However, the PAI LEE is still a determinant factor for colonization and persistence in other reservoirs (eg. Cattle) and its presence is more associated with pathogenic strains (Nataro and Kaper, 1998; Coombes *et al.*, 2011; Lewis *et al.*, 2015). Interestingly, the EHEC non-O157:H7 strains do not necessarily carry the LEE PAI. However, others colonization factors involved in the attachment, persistence and tissue tropism, from EHEC O157 and non-O157 have been well described, including Lpf, Ecp, Hcp, F9, curli, Type 1 fimbriae, Autotransporters (EhaA, EhaB), EspP, Saa, Cah (McWilliams and Torres, 2014; Monteiro *et al.*, 2016).

Besides the adhesion factors, toxins play a key role in EHEC pathogenesis and as mentioned above the Stx is one of the important virulence factors. This family of AB toxin is divided into Stx1 and Stx2 (with allelic variants); although both toxins or just one can be produced by the strain, Stx2 is considered one of the more potent toxins, thus more related to O157 infections. It has been showed that the use of antibiotics in the treatment against EHEC infections could lead to the cellular damage by increasing the production of this toxin activating the SOS system and by disrupting the bacterial membrane, causing release of the toxin (Kimmitt *et al.*, 2000).

The resulting secretion of the toxin into the blood stream, could lead to an even worsening of the disease (Pacheco and Sperandio, 2012). The Stx toxin is one of the factors involved in the hemolytic uremic syndrome (HUS) development, characterized by a triad of microangiopathic hemolytic anemia, thrombocytopenia, and acute renal failure, with the further chances of HUS development of 10% in children less than 10 years. Besides to trigger the HUS, and later renal failure, it is also responsible for strokes (Tarr *et al.*, 2005).
Due the implications of the Stx toxin during the course of the infection and restrictions in the treatment, different preventive strategies, as vaccines, have been implemented and are exemplified in the follow paragraphs.

Development of Stx-based treatments

Neutralizing the effect of the toxin has been one of the strategies to assuage its effect. Antibodies cαStx1 and cαStx2 have been directly engineered against the B subunit of Stx1 and A subunit of Stx2 respectively (Bitzan *et al.*, 2009). Evaluation of the tolerability and pharmakinetic profile has been investigated using chimeric anti-Stx1 and anti-Stx2 antibodies. This combination comprises the variable regions of the murine Stx1-neutralizing or Stx2-neutralizing monoclonal antibodies (mAbs) 13C4 or 11E10 respectively fused to human kappa light chain constant domain sequence and human immunoglobulin G1 (IgG1) heavy chain constant-domain sequence. These antibodies showed to be well tolerated and safe as antitoxins in healthy human volunteers in a clinical trial single dose study. Although, it has been showed that these mAbs are able to neutralizing the effect of the Stx toxins in mice experiments, it remains to elucidate whether they are able to avoid the development of HUS after diarrhea by EHEC (Bitzan *et al.*, 2009).
More recently, camelid single antibodies against Stx2 have been tested for their protective characteristics in against Stx2 (Mejias *et al.*, 2016). The anti-single chain antibodies (VHH) were obtained with two copies of anti-Stx2B VHH and one anti-sero albumin VHH. This trivalent molecule, administrated to mice, was able to decrease the toxicity in Stx2 lethal mouse model. Because the antitoxin effect of VHH, it is proposed as an alternative for treatment of assuage the HUS consequences (Mejias *et al.*, 2016).

Attenuated and Ghosts bacteria platforms

The gene regulators (global or specific) are important key players for the expression –or silencing - of virulence factors in pathogens. In fact, deleting those that promotes the expression

of specific virulence factors could be useful strategy to attenuate a pathogen. This approach has been applied to EHEC by deleting the LEE-encoded regulator (Ler). Ler is an important specific regulator that positively controls the expression of LEE genes involved in the A/E phenotype, as well as genes outside of the PAI, including adhesins and genes in the plasmid pO157. An EHEC O157:H7 86-24 strain deleted in *ler* and *stx*2 genes (ler/Stx2) but carrying a plasmid that expresses a detoxified version of Stx1A subunit and Stx2 A subunit was used as an attenuated vaccine candidate. These bacteria with a reduced toxicity and safe for animal administration were used to immunize mice via IP. Animals were challenged using EHEC O157:H7 wild type strain and after 6 days the wild type strain was not detected in faces, indicating a capability of the attenuated strain to reduce the wild type colonization. Also, a passive immunization was performed with suckling offspring born to mothers previously immunized with this vaccine; the results showed a 70% of surviving rate in passive immunized mice after being challenged with the wild-type EHEC bacteria (Liu *et al.*, 2009).

Further, EPEC is the closest pathotype to EHEC and both express the T3SS elements as EPEC-secreted proteins (Esps) proteins and intimin (important for the A/E phenotype). Using an EPEC, as live attenuated vaccine to assess the cross protection among the two pathotypes only partial protection against EHEC was shown. In fact, intragastric immunization with a clinical isolate EPEC O126:H6 and challenge using the EHEC O157:H7 wild-type strain, yielded sick mice but without dead reported. Interestingly, it was also shown that the EspB and intimin antibodies produced after the EPEC vaccination were cross-reactive against EspB (translocated protrein and effector that prevents phagocytosis) and intimin from EHEC (Calderon Toledo *et al.*, 2011).

Using a different delivery system such as *Salmonella* attenuated strain, recombinant EspA (300 amino acids of carboxyl-terminal), Intimin and the B subunit of Stx2 proteins have been expressed. Mice either only orally immunized or orally immunized and subcutaneous boosted were able to raise IgG and IgA specific antibodies against these three antigens at similar level. However, intimin specific antibodies levels, were higher in mice orally immunized and subcutaneous boosted than mice only orally immunized. Yet, specific IgG antibodies for Stx increased a week after the booster vaccination; IgA specific for Stx2B antibodies were only detected in feces and increased even more when boosted (Gu *et al.*, 2011).

Similarly, an attenuated bacteria *Salmonella* enterica serovar Typhimurium χ3987 (Δ*cya*, Δ*crp*, Δ*asd* and H683 Δ*aro* Δ*asd*) expressing heterolosgus proteins for EHEC as γ-intimin variant, encoded by the *eae* gene. The vaccination using these attenuated bacteria expressing the intimin was able to increase titers of IgG in serum and IgA in feces, indicating immune response from systemic and humoral immunity. The attenuated bacteria expressing the intimin was still

detectable in Peyer's patches and spleen while in feces the amount of CFU decreased from day 2 to 10 and later was kept it constant. Animal immunized with the attenuated *Salmonella* reduced the EHEC O157:H7 shedding post challenge and increasing the production of IgA and IgG (Oliveira *et al.*, 2012).

Also, a delivery antigen mucosal system used for the EHEC vaccine development was based on Bacillus Calmette-Guérin (BCG) (a live attenuated strain of *Mycobacterium bovis*), because its mucosal humoral immunereactivity. A recombinant rBCG-Stx2B, expressing the Stx2 subunit B, was generated and then used to immunize mice. This recombinant bacillus was able to increase the levels of Stx2 IgG in serum that were directly proportional to concentrations of CFU. The protection was confirmed when after challenge using an wild-type EHEC strain, there was a higher survival rate (63%) in immunized mice with higher CFU concentration than in immunized mice with a less CFU of rBCG-Stx2 or null rBCG (Fujii *et al.*, 2012).

Along attenuated bacteria strategies, a delivery system based on bacterial ghost (BG) platform for vaccine development against EHEC has been proposed. Bacterial ghosts of *E. coli* O157:H7 have been generated by the controlled expression of X174 lysis gene. This gene produces empty bacteria cell envelopes having the composition of cell surface as a living cell, such LPS, lipids, peptidoglycans (acting as adjuvants), lacking the capability of produce infection. These bacteria showed an antitoxicity effect on Vero cells culture and were safe for the administration in mice (Mayr *et al.*, 2005; Cai *et al.*, 2010). Animals orally immunized twice (days 0 and 28) with BG and then challenged (at day 55) stopped the shedding of the bacteria after day 3 post-EHEC and showed a rate survival of 93% (Cai *et al.*, 2010). In addition, rectal immunization showed a 100% of survival rate protecting mice against EHEC O157:H7 after challenge with bacterial shedding until 3 to 5 days after (Mayr *et al.*, 2012).

However, mice that were orally or intragastric immunized still showed disease symptoms from day 2 to 7 post-challenge including anorexia, slowing of activity, no stimulus reaction, and convulsions before death (Cai *et al.*, 2010). Data also showed that survived mice recovered after 7 days and dead mice developed glomerulus necrosis and enterocyte effacing. Interestingly, specific IgA and IgG antibodies titers were raised in sera and colon from mice immunized twice via oral with the BG (Cai *et al.*, 2010). In fact, the vaccination of mice with GB is, per se, able to trigger the immune response as it is observed in Th1/Th2 cell proliferation, higher INF levels in spleen cells, leading to increase titers of IgG and IgA, in both serum and colon samples (Mayr *et al.*, 2005; Cai *et al.*, 2010; Mayr *et al.*, 2012).

Most recently a BG (rSOBG), expressing a Stx chimerical protein composed by the Stx2A and Stx1B subunits (Stx2Am-Stx1B has been engineered. This rSOBG showed specific IgG and

IgA antibodies titer to StxA1 and StxB2, and the rate of animal survival was higher (52%) than native bacterial ghost-OBGs (12%) when challenged intragastrically with high dose of viable *E. coli* O157:H7. Also, the tissue damage on liver kidney and intestine from rSOBG immunized mice was not observed (Cai *et al.*, 2015).

Protein-based vaccines

Among several strategies applied for vaccine development against EHEC chimerical protein construction resulted attractive in the last years.

In this regard, a fusion protein between Stx2A and the N-terminus of EspA has been tested for its immunoreactivity. Mice subcutaneously immunized with EspA-StxA1 fusion protein showed high titers of specific IgG antibodies to EspA-StxA1. This humoral immuno-response, resulted in > 95% of mice survival after a challenge with crude toxin Stx2. Although *in-vitro* assays on HeLa cells showed that the anti-EspA-Stx2A1 serum was able to neutralize the action for Stx2, it did not prevent the adherence of bacteria to HeLa cells (Cheng *et al.*, 2009).

Another protein fusion constructed with the subunit B from two Shiga-toxins Stx1/Stx2 and a truncated intimin protein (SSI) was able to increase specific IgG antibody titers in mice. The immunized mice and orally challenged with EHEC O157:H7 88321 showed a 100% of survival rate. However, the protection using the chimerical vaccines is dependable of number of immunizations and the bacterial challenge dose (Cai *et al.*, 2011). The immunization with this SSI chimerical protein showed to avoid the pathological damage in colon and kidney tissues generating antibodies with anti-toxin and anti-adhesion effect, which was absent in the vaccines using the single proteins (Gansheroff *et al.*, 1999) (Gao *et al.*, 2009). Even though the toxins could have contributed with an adjuvant and neutralization effect, the fusion protein was not capable to avoid the wild-type adhesion in *in vitro* assays as in other subunits vaccines (Gansheroff *et al.*, 1999). Further, a vaccine composed by a Stx1B subunit and enzyme-inactive Stx2A subunit (Stx2Am-Stx1B, SAmB) induced Th2-mediated humoral immune response and its typical cytokines, IL4 and IL 10, but low level of INF-γ. Mice immunized with this chimerical protein and challenged with a lysed EHEC 88321 preparation showed 93% of survival, and even higher rates of survival were obtained challenging mice with the Stx1, Stx2 or Stx1/Stx2. However, disease manifestations were still evident (Gao *et al.*, 2011).

Also, specific peptides have been designed for the protection against EHEC. An example is the C terminal region of intimin associated with the AE. Antibodies obtained from vaccination with this fragment were able to reduce the bacterial attachment to Hep-2 cells cultures *in vitro* and

was also associated with protection in mice infected with *E. coli* O157:H7 (Wan *et al.*, 2011). B-cell epitopes of this protein were predicted by structural and antigenicity analysis, and proposed as synthetic vaccine candidates for EHEC (Wan *et al.*, 2011). A promising peptide, KT-12 (KASITEIKADKT) conjugated with KLH, was used to immunized mice either subcutanea (SC) or intranasal (IN) via. Both routes of immunization were able to induce high concentrations of IgG but higher in SC. By contrast, IgA titer was higher in IN immunization. Although this peptide did not show full protection to mice infected with the bacteria, it triggered the immune response in the animals (Zhang *et al.*, 2011).

Plant based vaccines

The safety aspect in vaccine development is one the important issues to take into consideration. An approach to reduce the risk of undesired side effect is the use plant-based vaccines targeting the mucosal immunity. The rational for using the plant cells is based on the idea of protecting the antigens from the protease degradation in the gastrointestinal tract by plants microencapsulation system, safe oral delivery and with a low production cost (Wen *et al.*, 2006; Amani *et al.*, 2009). An example was provided by a Stx toxoid generated by inactivating the toxin subunit A and expressed in *Nicotiana tabacum* (tobacco) NT-1 cell line. Mice immunized either, orally by feeding them with these cells expressing the toxoid, or by parenteral immunization and boosted orally, showed elevated specific Stx2 anti-IgA in fecal samples but higher in the oral immunized mice. Furthermore, sera of immunized mice were able to neutralize the Stx2 toxicity in Vero cells culture, but with higher neutralization titer using the sera from oral immunized mice (Wen *et al.*, 2006).
Another example of synthetic genes from EHEC was the chimerical gene composed by, *espA*, *eae*, and *tir* antigens (EIT). This gene, was codon optimized for expression in plant cells and cloned into a plant-expression vector, using CaMV35S (Cauliflower Mosaic Virus 35S) under the control of FAE promoters for tobacco and canola plants. Besides that, the solubility of the expressed protein was low but still, although a portion of the protein was expressed as soluble. The EIT protein was used to immunize mice either subcutaneously or orally, and later challenged with *E. coli* O157:H7, resulting in a significant reduction of the bacterial shedding. These immunized mice rose the levels of anti-EIT IgG and IgA and with a reduction of the bacterial colonization and histological damage (Amani *et al.*, 2009; Amani *et al.*, 2010; Amani *et al.*, 2011).

DNA vaccines

DNA vaccines have been used to avoid the use pathogens or bacterial traces that could yield in the development of disease in vaccine receiver. A DNA vaccine construct encoding the entire StxB2 subunit plus the last 32 amino acid residues of StxA2 was generated and cloned into the pGMS-CSF plasmid encoding the gene for murine granulocyte-macrophage colony-stimulating factor (GMS-CSF). The Stx2AΔAB DNA vaccine expressing the nontoxic Stx2 mutated form was able to increase IgG antibody titers and also conferred protection to immunized mice challenged with native Stx2. In addition, antibodies raised in mice conferred toxin neutralization in Vero culture cells (Bentancor *et al.*, 2009).

More recently, a selection of prospective DNA vaccine candidates was performed by bioinformatics analysis of EHEC O157:H7: EDL933 and Sakai strain genomes. The vaccine selection included, among others a putative pilin subunit gene (Z1538), the gene of a T3SS structural protein (*escC*), the C-terminal side of *escC*, and the gene encoding for an outer membrane protein (*iomW*) (Garcia-Angulo *et al.*, 2014; Tapia *et al.*, 2016). Mice were immunized via intranasal, and then challenged with the wild-type bacteria. In comparison to the entire *escC* gene, only its C-terminal portion resulted in a higher reduction of bacterial counts in feces, colon and caecum and also by triggering IgGs in sera and IgA in feces. The most interesting finding of this study was that the efficacy and immunoprotection of a vaccine candidate depends on its length and how it is presented to the immune system. In fact, the immunoprotection against EHEC was improved when only the C-terminal domain of EscC was used (Garcia-Angulo *et al.*, 2014; Tapia *et al.*, 2016).

Another DNA vaccine candidate used was the lymphocyte inhibitory factor A-/EHEC factor for adherence-1 gene (*lifA*/*efaA*). This gene encodes a toxin of 360kDa, mainly found in non-O157 EHEC strains and associated to LEE strains. It has been showed to be exposed on the surface of the EPEC bacteria, and possibly having a role in the colonization and adhesion by mucosal immunity regulation. This gene was originally found as a truncated form in EHEC O157:H7 EDL933 and annotated as *efa-1'*. EHEC carrying the truncated form showed a reduced adhesion to human colon cells, pointing out that Efa-1' protein has still a role in adhesion. Furthermore, mice vaccinated with this *efa-1'* showed IgM, IgG and IgA antibody titers. Intrasansal immunization using pVAX-efa1 showed higher levels of antigen specific mucosal IgA in nasal and broncheoalveolar lavages and in animal challenge pVAXefa-1 was also able to reduce the EHEC colonization in mice. (Riquelme-Neira *et al.*, 2015).

Polysaccharide-based vaccines

Polysaccharides in conjugate vaccine against Haemophilus influenzae type b, pneumococcal and meningococcal bacteraemia, and meningitis have been successfully used for vaccine development. *E. coli* isolates produce two serotype-specific surface polysaccharides namely the lipopolysaccharide (LPS) O antigen and the capsular polysaccharide K antigen. Variations in structures of these polysaccharides give rise to ~170 different O antigens and ~80 K antigens (Whitfield, 2006). Immunization with an O-specific polysaccharide of *E. coli* O157:H7, showed a significant increase of IgG against LPS. *E. coli* O157:H7 O-specific polysaccharide conjugated to recombinant exotoxin A of *P. aeruginosa* (O157-rEPA) administrated to 2-5-yeard-old children showed that there was an increase of >4 fold of IgG in serum after first week of immunization. At week 6, IgG anti-LPS increased in serum of >8 fold until the 20-folds higher at week 26, when actually there was not a difference among the groups receiving one dose or two doses, but >4-fold higher than the pre-immune sera. The serum samples had an antibacterial activity correlated to the IgG anti-LPS antibodies titers. More importantly, vaccinated children showed from mild to non-collateral reactions to the vaccine. This prospective vaccine showed to be a good candidate for its safeness and immunereactivity (Konadu *et al.*, 1994; Konadu *et al.*, 1998; Konadu *et al.*, 1999; Ahmed *et al.*, 2006).

Improving adjuvant effect

The suitability of an adjuvant is an important key in the vaccine development to enhance the immunogenicity and reactivity of the antigens on the host immune system. The vaccination with EspB or the C-terminal of γ-Intimin (280 amino acids) co-administrated with the MALP-2 adjuvant (TLR6 agonist) has showed to enhance the IgG specific antibodies after first intranasal immunization, in contrast the vaccination without MALP-2 increased the IgG titers after the second boost (Cataldi *et al.*, 2008).

The adjuvant effect has been also implemented by the combined expression of toxins and antigens at the same time, as a specific antigen combination. An example of this method was provided by the generation of a chimerical protein fusing Tir, Stx1B, Stx2B and the Zonula Occludens Toxin (Zot). Intranasal immunized mice showed higher IgA and IgG response against the chimerical protein than subcutaneous immunized mice and a reduced bacterial shedding in faces post EHEC challenge. The comparison of the adjuvant effect of chimerical protein with or without Zot has showed a higher protection level in Zot presence even though there were not IgG specific titer against it (Zhang *et al.*, 2011).

Future of the EHEC vaccine development

EHEC infections and associated diseases are also related to other 6 STEC serotypes, O26:H11, O45:H2, O103:H11, O11, O121:H19, O145, other than the canonical serotypes O113:H21 and O157:H7. The treatments and infection preventions, like the vaccine development, should also target these strains. More recently new approaches have been implemented to find new antigens; for example, an immunoproteomics analysis developed by a Chilean group, let them to detect antigens with immune reactive effect in patients infected with any of these STEC serotypes. They identified mainly outer membrane proteins, like OmpT and Cah that were immune reactive with sera from HUS patients. The genes encoding for these proteins are wide represented among *E. coli* pathotypes but also in commensal strains. Other detected proteins included FliC, Ag43 (ETEC), NmpC, OmpF, OmpC, OmpA, Hek, EF-Tu, and L-Asparaginase II. (Montero *et al.*, 2014).

ETEC THE MOST STUDIED INPEC FOR VACCINE DEVELOPMENT

Enterotoxigenic *E. coli*, is responsible for infection for traveler's diarrhea around the world, up to date it is been one of the most studied InPEC for vaccine targets. Developing countries are the most affected by ETEC infections and diarrheal diseases, being children the most concerned population. It was estimated that 200 million diarrheal cases and between 170,000 to 380,000 deaths occur annually (Isidean *et al.*, 2011; Chakraborty *et al.*, 2015).

The symptoms of an ETEC infection include dysentery, headache, fever and vomiting. The infection can last up to 5 days, without specific treatment or antibiotics, but the lethal cases are associated to kids because lack of immune protection and dehydration (Vidal *et al.*, 2016). ETEC is transmitted by consuming contaminated water, food, and person to person, being the poor sanitation one of the main factors in the pervasiveness of this pathogen in developing countries.

The main virulence factors associated to the pathogenesis of the infection are the colonization factors antigens (CFA) required for the ETEC colonization and establishment in the gut and the heat-labile (LT) and a heat-stable (ST) toxins that are responsible of water and electrolyte discharge during the infection (Vidal *et al.*, 2016). Regarding the CF, they are mainly referred as *E. coli* surface antigens (CS) composed by 26 characterized factors (CFA/I, CFA/III, CS2-

CS26). These factors are not present at the same time in the ETEC strains, but can be carried in different combinations. Nonetheless, the presence of other non-fimbrial adhesins as EtpA, Tia, TibA, TleA, and EaeH, expressed in the prototypic strain H10407, has been described (Vidal *et al.*, 2016). Concerning ETEC classification, the most common serotypes associated to ETEC are O6:H16 (LT/ST), O8:H9 (ST only), O25:NM (LT only), O78:H12 (ST only), O148:H28 (ST only), O153:H45 (ST only) and O169:H41 (ST only). However, serotypes isolated from different outbreaks do not necessarily belong to these serotypes or can also differ in their prevalence. Up to date, ETEC isolated can be gathered in 42 different clonal groups with a singular combination of CF and toxins (Croxen *et al.*, 2013). Because ETEC is still a public health problem in developing countries, there is a big effort to develop a vaccine to prevent and reduce the incidence of infections caused by this pathotype.

Toxin Based vaccines

The transcutaneous route has been used to immunize human volunteers with LT toxin delivered in skin patch, a strategy aimed to avoid toxic effect of the LT allowing antigen transfer from skin to Langerhans cells to lymph nodes. In these studies, designed as a double-blind, placebo-controlled trial, human volunteers were immunized by applying the patches on the skin arms then challenged with a LT/ETEC strain (McKenzie *et al.*, 2007; Frech *et al.*, 2008; Frerichs *et al.*, 2008). As in previous reports, in 100-97% of cases, there was an increased amount of anti-LT IgG and IgA (4-fold increase). Although, there was a delay in the onset of disease but also the need of intravenous fluids for volunteers, this strategy did not prevent the illness after challenge (McKenzie *et al.*, 2007; Frech *et al.*, 2008). The later refinement of the scale of disease parameters and correlation between symptoms and signs after human challenge let classify these specific patches vaccines with not efficacy (Porter *et al.*, 2016).

The STa is known to be poor immunogenic as antigen for vaccine use, however, its immunogenicity could be enhanced by being combined to stronger antigens, like LT or other adjuvants, and reducing its toxicity by changing essential amino acids. Mice immunized with a STa_{P13F}-LT_{R192G} toxoid fusion protein were able to develop IgG specific antibodies for LT and STa proteins in serum and feces and only IgA in feces. In addition, the STa and LT antibodies from fecal samples could reduce the cGMP and cAMP, respectively, in T-84 culture cells. This study indicates that the STaP13F toxoid is immunogenic when fused to LT toxoid and elicited neutralizing antitoxin antibodies. These findings will be useful for developing safe and effective toxoid vaccines linked to ETEC-STa-producing strains (Liu *et al.*, 2011).

Autotransporters

Immunoproteomic analysis has been also employed as strategy for the identification of antigens able to generate an immune response in the host during the process of infection. For this type of approach, reacting human or mice sera obtained after ETEC infections let the identification of immune reactive molecules in culture supernatant, outer membrane, and outer membrane vesicle preparations by matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS). In this analysis, hypothetical proteins homologous to other pathovars (UPEC, APEC, Crohn's disease associated isolates) and pathogens (*Vibrio cholera* and group A *streptococci*) were identified, suggesting their role in the pathogenesis. In addition, autotransporters (AT) such as EtpA, Antigen 43 and TibA, were also detected to be reactive during the infection process (Roy *et al.*, 2010).

In silico tools, have been used to the identification of other remarkable virulence factors as immunogenic molecules. From those studies, autotrasporters, important in biofilm formation, such as Autotransporter (pAT), Antigen 43 and EatA have been identified exclusively in ETEC or other pathotypes, but absent in commensal strains. Recombinant Ag43 and pAT autotransporters, were found capable of increasing fecal IgA and provide relative protection against ETEC intestinal colonization in immunized mice. Interestingly, these autotrasporters are also recognized by sera from patients with ETEC diarrhea, confirming their expression during ETEC infections (Harris *et al.*, 2011).

Adhesins

Adhesins have been wide employed as targets vaccines as several of them are considered virulence factors having an important role in colonization and bacterial pathogenesis.

The ETEC two-partner secretion protein A (EtpA) is an encoded plasmid protein and secreted non-fimbrial adhesin. When mice were immunized with the recombinant fully glycosylated form of this protein, EtpA inhibited the colonization of ETEC wilt-type strains. Also, the EtpA antibodies present in the sera, from mice immunized with the glycoprotein and enhanced by flagellin, were able to inhibit the colonization in *in vitro* assays (Roy *et al.*, 2009). In addition, a study of EtpA and CexE in OMV-flagelling free showed that these antigens were able to reduce the ETEC wild-type colonization in mice without the flagellin contribution (Roy *et al.*, 2011).

Attenuated bacteria platforms

Some studies have shown that strain ETEC E1392/75-2A is able to confer protection in 75% of human volunteers whom were immunized and challenged with ETEC wild-type strains. However, this strain induced mild diarrhea in 15% of volunteers. By deleting different genes onto the chromosomes, was possible to attenuate the strain and also to reduce its reactogenicity. Two different mutant strains were generated and attenuated by mutating aroC, ompR genes (PTL-002 mutant) and aroC, ompC and ompF genes (PTL-003 mutant). To be notice, both of these mutants express CFA/II: CS1, CS3 and are ST/LT spontaneously mutated. At first, these mutants were tested in a mouse model, and further tested in human volunteers, in who the reduction of adverse effects was observed. Interestingly, the strain PTL-003 showed higher immunogenicity level provided by increased IgA and IgG level and antibody response to CS1 and CS3. However, a poor protection in subjects orally immunized was obtained and only able to assuage diarrhea manifestations after the challenge with wild-type ETEC E24377A (Turner *et al.*, 2001; McKenzie *et al.*, 2006; McKenzie *et al.*, 2008).

Yet, the MEV or Etvax vaccine is another cocktail of four inactivated recombinant bacteria expressing separately different CFs in which CS6 expressed in a K12 strain and CFA/I, CS3, CS5 in ETEC O78 toxin-negative strain. The formalin or phenol attenuated bacteria were used in preclinical and clinical phase I studies (Lundgren *et al.*, 2014) and immunized altogether with a LCTBA hybrid protein (which has seven amino acids in CTB replaced by corresponding amino acids of LTB) and a double mutant LT (dmLT, that has two replacements in the A subunit, LTR192G/L211A (Norton *et al.*, 2011), (eliminating the toxic effect and leaving the adjuvant effect of the LT) (Lundgren *et al.*, 2014). In pre-clinical studies, they showed immunoreactivity in mice, resulting in high titers of fecal, jejunal tissue extracts and serum IgA and IgG enhanced the by the presence of the dmLT (Holmgren *et al.*, 2013) . Furthermore, a double-blind, randomized, placebo-controlled phase I, confirmed increasing titers of IgA and IgG in sera and fecal samples in humans. Volunteers orally immunized with the same vaccine but different amount of dmLT (10 µg or 25 µg) showed significant increase of intestinal, fecal and serum IgAs. Also, humans immunized with the formulation using 10 µg of dmLT showed significant response to all the antigens but even higher response to the CS6. What is more the vaccine combination was well tolerated and evaluated safe for human use. (Holmgren *et al.*, 2013; Lundgren *et al.*, 2014).

The attenuated bacterial strategies for vaccine development have also included the use of different bacterial species. In this matter, non-pathogenic and *Vibrio cholerae* strains were

engineered to express of CFA/I operon (cfaA-B-C-E), and killed by formalin to immunize mice via oral with CT as adjuvant. The results showed increased IgA and IgG titers in serum. The antigen delivery was thought to be favored by the fact that *V. cholerae* can attach to M cells, probably enhancing the bacterial attachment on the intestinal mucosal by the over expression of CFA/I. Increasing titers of IgA and IgG+M were showed against *V. cholerae* O1 LPS after immunization (Tobias *et al.*, 2008). In the same way, when CS2 was added to the formalin-killed *E. coli* K12 over-expressing CFA/I and orally administrated a similar immuneresponse (IgG+M increase in sera and IgA in fecal sample) against *cfaB* from CFA/I and *cotA* from CS2 was observed. (Tobias *et al.*, 2010).
Following the same strategy as the PTL attenuated vaccines, mentioned above, three different others ETEC strains were attenuated by the same gene mutations and used as prospective vaccines. These strains expressed diverse CFAs elements including CFA/I (ACAM2010 strain), CS2 and CS3 (ACAM2007 strain) and CS1, CS2, and CS3 (ACAM2017 strtrain). As showed by PTL-003, mentioned above, there was an immuneresponse against those CFAs elements increasing the IgA level in human volunteers orally immunized. However, neither IL-6 nor IL-8, inflammatory interleukins were detected (Daley *et al.*, 2007).
Further, the CS21, an adhesion factor belonging to the Class-B Type 4 pili as BFP and TCP, was tested as a potential vaccine candidate. The CS21 including its major subunit LngA was able to confer certain immunoreactivity in mice model. The immunogenicity of CS21 and LngA in different adjuvants formulations showed that CS21/LngA combined with cholera Toxin and administrated via intraperitoneal, increased specific IgG antibodies in serum and IgA antibody levels in fecal and intestinal lavages. Interestingly, anti-LngA antibodies were also raised in an intraperitoneal administration of CS21 formulated with incomplete Freund's adjuvant and after mice challenge, this formulation was also able to stop the bacteria shedding. By contrast, CS21 plus CT as adjuvant was more effective when was intranasal administrated. The most relevant aspect from this study was that supports the idea of using CF as promising antigens and showing that the administration routes play an important role to elicit an effective immuneresponse and protection (Zhang *et al.*, 2016).
Other colonization factors from different ETEC strains have been studied as vaccine candidates, mainly because there is a usual combination of co-expression of these factors in ETEC strains, making it as a hallmark to distinguish the different strains. To cover a wide range of the ETEC strains variability involved in diarrheal infections, a non-toxigenic *E. coli* was modified to express every CS2, CS4, CS5, or CS6 independently or co-expressed with CFA/I. Mice immunization with the recombinant bacteria, co-expressing CFA/I and CS2 induced the

IgG+IgM antibodies in sera and IgA in fecal samples. The antibodies titers for the recombinant bacteria were higher than wild-type stains. The CS6 overexpressed in a non-toxigenic strain, and used as killed vaccine, increased the IgG+IgM antibodies in sera and fecal IgA antibodies like the other strains (Tobias *et al.*, 2010; Tobias *et al.*, 2011). Although CS6 was previously reported as a weak antigen to induce the immune response, this study was capable to improve CS6 properties as a vaccine candidate.

More recently, the generation of three strains comprised in the ACE527 ETEC complex, has been the most promising vaccine in phase II studies. The strains in the ACE527 have been attenuated, converted to antibiotic sensible, and genetically engineered to overexpressing different colonization factors. All of them have been attenuated by mutation on *aroC, ompC and ompF*, showing safeness and being immunogenic in humans as PTL-003 attenuated strain (Turner *et al.*, 2001). These three strains ACAM2022 (O141:H5, expressing CS5 and CS6), ACAM2025 (O39:H12, expressing CFA/I) and ACAM2027 (O71:H-, expressing CS2, CS3, and CS1) express also the heat labile toxin pentamer B subunits. In particular, the LTB gene encoding for heat labile toxin pentamer B was inserted into ACAM2022 genome to assure its stability and avoid their lost during the vaccination. In addition, the gene encoding for CS1 was inserted into ACAM2027 genome by replacing *ompC*. Although other CFs as CS3, CS5m CS6 and CFA/I were maintained into their native plasmid, those were showed to be stable (Turner *et al.*, 2011).

The human clinical trials, testing two different oral doses 3 weeks apart of ACE527 (doses of 10^{11} or 3 X 10^{10} CFU each strain) in a CeraVacx buffer, showed that IgA and IgG increased in serum, for LTB, CFA/I, CS3, CS5, and CS6. After the immunization, subjects were challenged with the ETEC H10407 wild-type strain. As a result, the vaccination helped to shortened the duration of the diarrhea and reducing the shedding of the wild-type strain, up to 20-fold less than control groups, 2 days after challenge, and conferring a protection level ranging between 33% to 98% against the wild type strain. Re-challenged subjects decreased even more the bacterial shedding (Harro *et al.*, 2011a; Harro *et al.*, 2011b). Overall, this ACE527 formulated in CeraVax showed to have a reduction of 29% of diarrhea and 26.5% of efficacy (Harro *et al.*, 2011a; Darsley *et al.*, 2012; Porter *et al.*, 2016). Furthermore, when ACE527 vaccine combination was formulated with LTR192GL211A adjuvant, the efficacy raised up to 50% (Porter *et al.*, 2016).

Outer Membrane Vesicles

The outer membrane vesicles as antigen delivery system, have been largely employed. The expression of heterologous surface antigens has been applied not only to amplify the protection against different pathogens but also to increase the antigenicity of such antigens. Recently, this system was used in two different ways. The first one used an ETEC mutated in the lipid A acyltranferase (*msbB*) and Labile toxin subunit A (*eltA*), (EΔ*msbB*Δ*eltA*) to decrease the OMV toxicity level. The OMV generated from this mutant were used to immunize adult mice and test the immunoreactivity. It was shown that the detoxified OMV yielded higher titers of IgG1, IgM, and IgA, in comparison to the OMV from a wild-type strain. Later, offspring, born to immunized mothers with the OMV- EΔ*msbB*Δ*eltA*, passively immunized by breast feeding and challenged with ETEC, displayed low colonization level. In the second part of this approach, ETEC antigens including the adhesins FliC and CFA/I, were expressed as heterologous proteins in *Vibrio cholera* OMV. This strain was engineered by deletion of *msbB*, cholera toxin subunits A and B (*ctxAB)* and the major flagelling FlaA (*flaA*) (VΔ*msbB*Δ*ctxAB*Δ*flaA*). As in the previous test with the OMV from ETEC EΔ*msbB*Δ*eltA*, the immuneresponse against VΔ*msbB*Δ*ctxAB*Δ*flaA* expressing FliC-CFA/I, displayed the same pattern in adult mice and a similar colonization level in the neonatal mice model. While an immuneresponse against both heterologous expressed ETEC antigens in *Vibrio cholerae* was induced, none of these approaches resulted in an improved protection level (Leitner *et al.*, 2015).

IMMUNOPROTECTION FOR EPEC INFECTIONS

Enteropathogenic *E. coli*, non-invasive bacteria, are able to colonize small intestine, causing moderate to acute diarrhea mainly in less than 2 years old children in developing countries. The peculiar characteristic of this pathogen and pathotype is the presence of the LEE pathogenic genomic island that encodes virulence factors, such as intimin and the translocated intimin receptor (Tir), associated with the T3SS necessary to produce attaching and effacing (A/E) lesions on the intestinal microvilli. The A/E lesion consists in the rearrangement of the actin and tight attachment of the bacteria to the host cells by the translocation of effectors. Tir, that serves as receptor for intimin and promotes the actin rearrangement, is one of the main protein translocated from the bacterium to the eukaryotic cell through the T3SS. In contrast to EHEC LEE-positive, intimin and Tir are important for the adhesion and establishment of EPEC to the eukaryotic cells. Also, it is well known the EPEC's capability of a localized adherence (LA) produced by a bundle-forming pili (BFP, being BfpA its major subunit); this pili is encoded by the plasmid *E. coli* adherence factor (pEAE) (Ochoa *et al.*, 2008). Nonetheless, this pathotype

has been subcategorized in: i) Typical EPEC (tEPEC) carrying the EPEC adherence factor plasmid (pEAF) and LEE island and ii) Atypical EPEC (aEPEC) which lack the pEAF (Croxen *et al.*, 2013; Gomes *et al.*, 2016; Scaletsky and Fagundes-Neto, 2016). In the case of the aEPEC, that lacks the BFP, it has been showed that they can produce a LA-like pattern promoted by the intimin (subtype omicron), and in some cases the aggregative or diffuse patters (Hernandes *et al.*, 2008; Gomes *et al.*, 2016).

The most common O serogroups for the classical EPEC are: O26, O39, O55, O86, O88, O103, O111, O114, O119, O125ac, O126, O127, O128ab, O142, O145, O157 and O158. In the other hand, even that the most recurrent aEPEC serogroups are O51, O145, O26, O55, and O111, many others of the aEPECs do not belong to the tEPEC same serogroups and some are neither O nor H typeable (Hernandes *et al.*, 2009; Hu and Torres, 2015). The tEPEC is most commonly recovered from humans while aEPECs are recovery from different farmer and domestic animals considered as a zoonotic pathogen (Hernandes *et al.*, 2009; Gomes *et al.*, 2016). However, latest reports have showed an increased emergence of aEPEC in developed and developing countries (Ochoa and Contreras, 2011; Ingle *et al.*, 2016).

The incidence and prevalence numbers of EPEC infections are variable according to the diagnosis methods, in fact the molecular methods targeting specific genes (as intimin) seems to show an incidence of 5-10% of pediatric diarrheal cases in developing countries whereas diagnosis made on Hep-2 adherence-pattern and serotyping has shown an average prevalence rates from 10 to 20% (Ochoa *et al.*, 2008). More recently it has been shown that aEPEC have a major prevalence in both developing and developed countries than tEPEC (Ochoa *et al.*, 2008).

Passive immunization against EPEC

A natural immune protection after an initial EPEC infection has been shown and antibodies can protect against future infections. Breast feeding has an important role in the natural immune protection against EPEC infections. Studies of colostra from mothers in developing countries have shown that IgA antibodies, mainly against intimin, EspA, EspB, EspD, EspC and BFP, can be transmitted by colostra. These antibodies can prime and protect the neonates in the first hours of birth against EPEC. (Parissi-Crivelli *et al.*, 2000; Noguera-Obenza *et al.*, 2003; Durand *et al.*, 2013). In addition, EspA, B, C, and D, immuneresponse might confer cross protection against EHEC strains that carry the LEE locus. The prevalence of IgA antibodies against BfpA and EspB has been detected also in stool specimens from breastfed-pediatric patients with acute diarrhea, but absent in healthy children and non-breastfed (Quintana Flores *et al.*, 2002). The

main prevalence of anti-BfpA fecal antibodies has shown the immogenicity of BfpA, however their functionality as blocking or neutralizing antibodies have not fully elucidated. (de Souza Campos Fernandes *et al.*, 2003).

For EPEC vaccine development a number of candidates have been considered. Among these, EspB is one of the most important proteins for the EPEC pathogenesis which contributes to the attaching and effacing lesion in epithelial cells, it helps the Tir phosphorylation, the entrance into the host cells and mediates anti-phagocytosis by inhibiting myosin function (Taylor *et al.*, 1998; Iizumi *et al.*, 2007). Thus, an EspB-EPEC mutant has been used in human immunization, showing that it was able to confer protection to 9 out 10 human volunteers, from whom Diarrhea was absent. The mutant strain was present in the stools in fewer rates than the wild-type strains and for the same period of time. In addition, a biopsy performed to two volunteers showed not destruction of microvillus brush border (Tacket *et al.*, 2000).

Besides all mention above, up to date there is still not any vaccine available to prevent the infections against EPEC, which occurs mainly in developing countries and where acute diarrhea due the infection of this pathogen is still a public health concern. The most recent information obtained both from the genome sequencing and the maternal/passive immunization will allow the scrutiny for novel antigens that will be used for vaccine development (Torres, 2017b).

CONCLUSIONS AND FUTURE PERSPECTIVES

The highly genomic plasticity of *E. coli* is one of the issues to develop an effective treatment against the disease assortment caused by this bacterium. *E. coli* strains share a common core genome which is highly conserved among all the strains while the encoding genes related to diverse pathogenic functions, involved in infection, disease development and host-pathogen interaction varies among phatotypes. The whole genome sequencing (WGS) has unraveled the variations of the bacterial clonal population, and opened new lights on the evolution of the pathogenic strains, respective to a non-pathogenic strain, to a pathotype prototype or to a specific serotype (Sahl *et al.*, 2015 Moriel *et al.*, 2012; Ingle *et al.*, 2016). In addition, the comparison of virulence factors presence among different *E. coli* phylogenetic groups and clinical outcomes (severe, lethal, non-lethal or asymptomatic diarrheal cases) suggest that the associated genes can be lost and reacquired many times. Moreover, the severity of the disease and the closes relationship among the isolates from the same and similar clinical outcomes might suggest that also the host has an important influence in the severity of the disease in addition to the genomic background of the bacteria. In the same way, infections can be

generated by clonal populations, hypervirulent clones and in some cases by multiple pathogenic strains in one individual, each one with specific and unique genomic and virulent characteristics (Sahl *et al.*, 2015; Hazen *et al.*, 2016).

Nonetheless, as illustrated in this review, different studies have led to the discovery of potential vaccine candidates related at least to certain or most common InPEC infections. A number of different approaches have been used for vaccine candidate identification. Nowadays, interpreting the data from "omics" approaches such as WGS, transcriptomic and proteomics, will empowered the new antigen discovery and improvement of existing prospective vaccines. The efforts to prevent diarreagenic diseases produced by intestinal pathogenic *E. coli* should be strength, not only for the ETEC or EHEC infections but to the other pathotypes that are becoming relevant agents in the public health. Importantly, treatments and preventive intervetion, should also target non-O157 EHECs serotypes, and other pathotypes as DAEC, EIEC, EAEC for which vaccines development and discovery studies are scarce or nulls (Bouzari *et al.*, 2010). Recently has been proposed that multidisciplinary research, collaboration and partnerships should integrate the One Health concept, as done in Latin America in recent years (Torres, 2017a). In particular, this concept states that the human health is connected to the relationship between environment, animals and human beings. For this matter, extending research to the animal pathogenic strains should be necessary. Examples or this are *E. coli* research studies in animal reservoirs (Etcheverría *et al.*, 2016).

The battle against InPEC infections is still ongoing and the perspectives for the vaccine development need a multidisciplinary research to get advances in the treatments and vaccine development. Additional efforts, not only from biomedical or basic research but also from clinical, veterinarian, and public health fields, will be required to better understand the disease proliferation in different settings and design new preventive strategies (CDC).

AUTHOR CONTRIBUTIONS

MR-L and RR, wrote the manuscript. All authors contributed to the preparation of the manuscript, to the ideas and concepts contained in it. All authors reviewed and approved the manuscript.

FUNDING

The study was funded by EU FP7 Marie Curie Actions ITN (Initiale Training Network) EID (European Industrial Doctorate) DIS*Co* (A multidisciplinary Doctoral Industrial School on novel preventive strategies against *E. Coli* infections; http://www.discoproject.eu/; n°FP7-PEOPLE-607611) project. Maricarmen ROJAS-LOPEZ and Ricardo MONTEIRO are Marie Curie PhD Research Fellow granted by ITN EID DIS*Co* (DISCo Full Partners: GSK/INRA; DISCo Associated Partner: ROMA TRE

University and Université Auvergne).

CONFLICT OF INTEREST STATEMENT

RR and MP are permanent employees of GSK. The authors declare that GSK provided support in the form of salaries. All the other authors declare that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

Ahmed, A., Li, J., Shiloach, Y., Robbins, J.B., and Szu, S.C. (2006). Safety and immunogenicity of Escherichia coli O157 O-specific polysaccharide conjugate vaccine in 2-5-year-old children. *J Infect Dis* 193(4), 515-521. doi: 10.1086/499821.

Amani, J., Mousavi, S.L., Rafati, S., and Salmanian, A.H. (2009). In silico analysis of chimeric espA, eae and tir fragments of Escherichia coli O157:H7 for oral immunogenic applications. *Theor Biol Med Model* 6, 28. doi: 10.1186/1742-4682-6-28.

Amani, J., Mousavi, S.L., Rafati, S., and Salmanian, A.H. (2011). Immunogenicity of a plant-derived edible chimeric EspA, Intimin and Tir of Escherichia coli O157:H7 in mice. *Plant Sci* 180(4), 620-627. doi: 10.1016/j.plantsci.2011.01.004.

Amani, J., Salmanian, A.H., Rafati, S., and Mousavi, S.L. (2010). Immunogenic properties of chimeric protein from espA, eae and tir genes of Escherichia coli O157:H7. *Vaccine* 28(42), 6923-6929. doi: 10.1016/j.vaccine.2010.07.061.

Bentancor, L.V., Bilen, M., Brando, R.J., Ramos, M.V., Ferreira, L.C., Ghiringhelli, P.D., *et al.* (2009). A DNA vaccine encoding the enterohemorragic Escherichia coli Shiga-like toxin 2 A2 and B subunits confers protective immunity to Shiga toxin challenge in the murine model. *Clin Vaccine Immunol* 16(5), 712-718. doi: 10.1128/CVI.00328-08.

Bitzan, M., Poole, R., Mehran, M., Sicard, E., Brockus, C., Thuning-Roberson, C., *et al.* (2009). Safety and pharmacokinetics of chimeric anti-Shiga toxin 1 and anti-Shiga toxin 2 monoclonal antibodies in healthy volunteers. *Antimicrob Agents Chemother* 53(7), 3081-3087. doi: 10.1128/AAC.01661-08.

Bouzari, S., Dashti, A., Jafari, A., and Oloomi, M. (2010). Immune response against adhesins of enteroaggregative Escherichia coli immunized by three different vaccination strategies (DNA/DNA, Protein/Protein, and DNA/Protein) in mice. *Comp Immunol Microbiol Infect Dis* 33(3), 215-225. doi: 10.1016/j.cimid.2008.10.002.

Brzuszkiewicz, E., Thurmer, A., Schuldes, J., Leimbach, A., Liesegang, H., Meyer, F.D., *et al.* (2011). Genome sequence analyses of two isolates from the recent Escherichia coli outbreak in Germany reveal the emergence of a new pathotype: Entero-Aggregative-Haemorrhagic Escherichia coli

(EAHEC). *Arch Microbiol* 193(12), 883-891. doi: 10.1007/s00203-011-0725-6.

Cai, K., Gao, X., Li, T., Hou, X., Wang, Q., Liu, H., *et al.* (2010). Intragastric immunization of mice with enterohemorrhagic Escherichia coli O157:H7 bacterial ghosts reduces mortality and shedding and induces a Th2-type dominated mixed immune response. *Can J Microbiol* 56(5), 389-398. doi: 10.1139/w10-025.

Cai, K., Gao, X., Li, T., Wang, Q., Hou, X., Tu, W., *et al.* (2011). Enhanced immunogenicity of a novel Stx2Am-Stx1B fusion protein in a mice model of enterohemorrhagic Escherichia coli O157:H7 infection. *Vaccine* 29(5), 946-952. doi: 10.1016/j.vaccine.2010.11.035.

Cai, K., Tu, W., Liu, Y., Li, T., and Wang, H. (2015). Novel fusion antigen displayed-bacterial ghosts vaccine candidate against infection of Escherichia coli O157:H7. *Sci Rep* 5, 17479. doi: 10.1038/srep17479.

Calderon Toledo, C., Arvidsson, I., and Karpman, D. (2011). Cross-reactive protection against enterohemorrhagic Escherichia coli infection by enteropathogenic E. coli in a mouse model. *Infect Immun* 79(6), 2224-2233. doi: 10.1128/IAI.01024-10.

Cataldi, A., Yevsa, T., Vilte, D.A., Schulze, K., Castro-Parodi, M., Larzabal, M., *et al.* (2008). Efficient immune responses against Intimin and EspB of enterohaemorragic Escherichia coli after intranasal vaccination using the TLR2/6 agonist MALP-2 as adjuvant. *Vaccine* 26(44), 5662-5667. doi: 10.1016/j.vaccine.2008.07.027.

CDC. Available: www.cdc.gov/onehealth/basics/index.html [Accessed].

Chakraborty, S., Harro, C., DeNearing, B., Ram, M., Feller, A., Cage, A., *et al.* (2015). Characterization of Mucosal Immune Responses to Enterotoxigenic Escherichia coli Vaccine Antigens in a Human Challenge Model: Response Profiles after Primary Infection and Homologous Rechallenge with Strain H10407. *Clin Vaccine Immunol* 23(1), 55-64. doi: 10.1128/CVI.00617-15.

Cheng, Y., Feng, Y., Luo, P., Gu, J., Yu, S., Zhang, W.J., *et al.* (2009). Fusion expression and immunogenicity of EHEC EspA-Stx2Al protein: implications for the vaccine development. *J Microbiol* 47(4), 498-505. doi: 10.1007/s12275-009-0116-8.

Coombes, B.K., Gilmour, M.W., and Goodman, C.D. (2011). The evolution of virulence in non-o157 shiga toxin-producing Escherichia coli. *Front Microbiol* 2, 90. doi: 10.3389/fmicb.2011.00090.

Croxen, M.A., and Finlay, B.B. (2010). Molecular mechanisms of Escherichia coli pathogenicity. *Nat Rev Microbiol* 8(1), 26-38. doi: 10.1038/nrmicro2265.

Croxen, M.A., Law, R.J., Scholz, R., Keeney, K.M., Wlodarska, M., and Finlay, B.B. (2013). Recent advances in understanding enteric pathogenic Escherichia coli. *Clin Microbiol Rev* 26(4), 822-880. doi: 10.1128/CMR.00022-13.

Daley, A., Randall, R., Darsley, M., Choudhry, N., Thomas, N., Sanderson, I.R., *et al.* (2007). Genetically modified enterotoxigenic Escherichia coli vaccines induce mucosal immune responses without inflammation. *Gut* 56(11), 1550-1556. doi: 10.1136/gut.2006.112805.

Darsley, M.J., Chakraborty, S., DeNearing, B., Sack, D.A., Feller, A., Buchwaldt, C., *et al.* (2012). The oral, live attenuated enterotoxigenic Escherichia coli vaccine ACE527 reduces the incidence and severity of diarrhea in a human challenge model of diarrheal disease. *Clin Vaccine Immunol* 19(12), 1921-1931. doi: 10.1128/CVI.00364-12.

De Gregorio, E., and Rappuoli, R. (2012). Vaccines for the future: learning from human immunology. *Microb Biotechnol* 5(2), 149-155. doi: 10.1111/j.1751-7915.2011.00276.x.

de Souza Campos Fernandes, R.C., Quintana Flores, V.M., Sousa de Macedo, Z., and Medina-Acosta, E. (2003). Coproantibodies to the enteropathogenic Escherichia coli vaccine candidates BfpA and EspB in breastfed and artificially fed children. *Vaccine* 21(15), 1725-1731. doi: S0264410X0200525X [pii].

Durand, D., Ochoa, T.J., Bellomo, S.M., Contreras, C.A., Bustamante, V.H., Ruiz, J., *et al.* (2013). Detection of secretory immunoglobulin A in human colostrum as mucosal immune response against proteins of the type III secretion system of Salmonella, Shigella and enteropathogenic Escherichia coli. *Pediatr Infect Dis J* 32(10), 1122-1126. doi: 10.1097/INF.0b013e318293306c.

Etcheverría, A.I., Lucchesi, P.M.A., Krüger, A., Bentancor, A.B., and Padola, N.L. (2016). "Escherichia coli in Animals," in *Escherichia coli in the Americas,* ed. A.G. Torres. (Cham: Springer International Publishing), 149-172.

Feavers, I.M., and Maiden, M.C.J. (2017). Recent Progress in the Prevention of Serogroup B Meningococcal Disease. *Clin Vaccine Immunol* 24(5). doi: 10.1128/CVI.00566-16.

Foster, M.A., Iqbal, J., Zhang, C., McHenry, R., Cleveland, B.E., Romero-Herazo, Y., *et al.* (2015). Enteropathogenic and enteroaggregative E. coli in stools of children with acute gastroenteritis in Davidson County, Tennessee. *Diagn Microbiol Infect Dis* 83(3), 319-324. doi: 10.1016/j.diagmicrobio.2015.07.016.

Frank, C., Werber, D., Cramer, J.P., Askar, M., Faber, M., an der Heiden, M., *et al.* (2011). Epidemic profile of Shiga-toxin-producing Escherichia coli O104:H4 outbreak in Germany. *N Engl J Med* 365(19), 1771-1780. doi: 10.1056/NEJMoa1106483.

Frech, S.A., Dupont, H.L., Bourgeois, A.L., McKenzie, R., Belkind-Gerson, J., Figueroa, J.F., *et al.* (2008). Use of a patch containing heat-labile toxin from Escherichia coli against travellers' diarrhoea: a phase II, randomised, double-blind, placebo-controlled field trial. *Lancet* 371(9629), 2019-2025. doi: 10.1016/S0140-6736(08)60839-9.

Frerichs, D.M., Ellingsworth, L.R., Frech, S.A., Flyer, D.C., Villar, C.P., Yu, J., *et al.* (2008). Controlled, single-step, stratum corneum disruption as a pretreatment for immunization via a patch. *Vaccine* 26(22), 2782-2787. doi: 10.1016/j.vaccine.2008.02.070.

Fujii, J., Naito, M., Yutsudo, T., Matsumoto, S., Heatherly, D.P., Yamada, T., *et al.* (2012). Protection by a recombinant Mycobacterium bovis Bacillus Calmette-Guerin vaccine expressing Shiga toxin 2 B subunit against Shiga toxin-producing Escherichia coli in mice. *Clin Vaccine Immunol* 19(12), 1932-1937. doi: 10.1128/CVI.00473-12.

Gansheroff, L.J., Wachtel, M.R., and O'Brien, A.D. (1999). Decreased adherence of enterohemorrhagic Escherichia coli to HEp-2 cells in the presence of antibodies that recognize the C-terminal region of intimin. *Infect Immun* 67(12)**,** 6409-6417.

Gao, X., Cai, K., Li, T., Wang, Q., Hou, X., Tian, R., *et al.* (2011). Novel fusion protein protects against adherence and toxicity of enterohemorrhagic Escherichia coli O157:H7 in mice. *Vaccine* 29(38)**,** 6656-6663. doi: 10.1016/j.vaccine.2011.06.106.

Gao, X., Cai, K., Shi, J., Liu, H., Hou, X., Tu, W., *et al.* (2009). Immunogenicity of a novel Stx2B-Stx1B fusion protein in a mice model of Enterohemorrhagic Escherichia coli O157:H7 infection. *Vaccine* 27(14)**,** 2070-2076. doi: 10.1016/j.vaccine.2009.01.115.

Garcia-Angulo, V.A., Kalita, A., Kalita, M., Lozano, L., and Torres, A.G. (2014). Comparative genomics and immunoinformatics approach for the identification of vaccine candidates for enterohemorrhagic Escherichia coli O157:H7. *Infect Immun* 82(5)**,** 2016-2026. doi: 10.1128/IAI.01437-13.

Giuliani, M.M., Adu-Bobie, J., Comanducci, M., Arico, B., Savino, S., Santini, L., *et al.* (2006). A universal vaccine for serogroup B meningococcus. *Proc Natl Acad Sci U S A* 103(29)**,** 10834-10839. doi: 10.1073/pnas.0603940103.

Goldwater, P.N., and Bettelheim, K.A. (2012). Treatment of enterohemorrhagic Escherichia coli (EHEC) infection and hemolytic uremic syndrome (HUS). *BMC Med* 10**,** 12. doi: 10.1186/1741-7015-10-12.

Gomes, T.A.T., Yamamoto, D., Vieira, M.A.M., and Hernandes, R.T. (2016). "Atypical Enteropathogenic Escherichia coli," in *Escherichia coli in the Americas,* ed. A.G. Torres. (Cham: Springer International Publishing), 77-96.

Gu, J., Ning, Y., Wang, H., Xiao, D., Tang, B., Luo, P., *et al.* (2011). Vaccination of attenuated EIS-producing Salmonella induces protective immunity against enterohemorrhagic Escherichia coli in mice. *Vaccine* 29(43)**,** 7395-7403. doi: 10.1016/j.vaccine.2011.07.069.

Harris, J.A., Roy, K., Woo-Rasberry, V., Hamilton, D.J., Kansal, R., Qadri, F., *et al.* (2011). Directed evaluation of enterotoxigenic Escherichia coli autotransporter proteins as putative vaccine candidates. *PLoS Negl Trop Dis* 5(12)**,** e1428. doi: 10.1371/journal.pntd.0001428.

Harro, C., Chakraborty, S., Feller, A., DeNearing, B., Cage, A., Ram, M., *et al.* (2011a). Refinement of a human challenge model for evaluation of enterotoxigenic Escherichia coli vaccines. *Clin Vaccine Immunol* 18(10)**,** 1719-1727. doi: 10.1128/CVI.05194-11.

Harro, C., Sack, D., Bourgeois, A.L., Walker, R., DeNearing, B., Feller, A., *et al.* (2011b). A combination vaccine consisting of three live attenuated enterotoxigenic Escherichia coli strains expressing a range of colonization factors and heat-labile toxin subunit B is well tolerated and immunogenic in a placebo-controlled double-blind phase I trial in healthy adults. *Clin Vaccine Immunol* 18(12)**,** 2118-2127. doi: 10.1128/CVI.05342-11.

Hazen, T.H., Donnenberg, M.S., Panchalingam, S., Antonio, M., Hossain, A., Mandomando, I., *et al.*

(2016). Genomic diversity of EPEC associated with clinical presentations of differing severity. *Nature Microbiology* 1**,** 15014. doi: 10.1038/nmicrobiol.2015.14
http://www.nature.com/articles/nmicrobiol201514 - supplementary-information.

Hernandes, R.T., Elias, W.P., Vieira, M.A., and Gomes, T.A. (2009). An overview of atypical enteropathogenic Escherichia coli. *FEMS Microbiol Lett* 297(2)**,** 137-149. doi: 10.1111/j.1574-6968.2009.01664.x.

Hernandes, R.T., Silva, R.M., Carneiro, S.M., Salvador, F.A., Fernandes, M.C., Padovan, A.C., *et al.* (2008). The localized adherence pattern of an atypical enteropathogenic Escherichia coli is mediated by intimin omicron and unexpectedly promotes HeLa cell invasion. *Cell Microbiol* 10(2)**,** 415-425. doi: 10.1111/j.1462-5822.2007.01054.x.

Holmgren, J., Bourgeois, L., Carlin, N., Clements, J., Gustafsson, B., Lundgren, A., *et al.* (2013). Development and preclinical evaluation of safety and immunogenicity of an oral ETEC vaccine containing inactivated E. coli bacteria overexpressing colonization factors CFA/I, CS3, CS5 and CS6 combined with a hybrid LT/CT B subunit antigen, administered alone and together with dmLT adjuvant. *Vaccine* 31(20)**,** 2457-2464. doi: 10.1016/j.vaccine.2013.03.027.

Hu, J., and Torres, A.G. (2015). Enteropathogenic Escherichia coli: foe or innocent bystander? *Clin Microbiol Infect* 21(8)**,** 729-734. doi: 10.1016/j.cmi.2015.01.015.

Iizumi, Y., Sagara, H., Kabe, Y., Azuma, M., Kume, K., Ogawa, M., *et al.* (2007). The enteropathogenic E. coli effector EspB facilitates microvillus effacing and antiphagocytosis by inhibiting myosin function. *Cell Host Microbe* 2(6)**,** 383-392. doi: 10.1016/j.chom.2007.09.012.

Ingle, D.J., Tauschek, M., Edwards, D.J., Hocking, D.M., Pickard, D.J., Azzopardi, K.I., *et al.* (2016). Evolution of atypical enteropathogenic E. coli by repeated acquisition of LEE pathogenicity island variants. *Nature Microbiology* 1**,** 15010. doi: 10.1038/nmicrobiol.2015.10
http://www.nature.com/articles/nmicrobiol201510 - supplementary-information.

Isidean, S.D., Riddle, M.S., Savarino, S.J., and Porter, C.K. (2011). A systematic review of ETEC epidemiology focusing on colonization factor and toxin expression. *Vaccine* 29(37)**,** 6167-6178. doi: 10.1016/j.vaccine.2011.06.084.

Kaper, J.B., Nataro, J.P., and Mobley, H.L. (2004). Pathogenic Escherichia coli. *Nat Rev Microbiol* 2(2)**,** 123-140. doi: 10.1038/nrmicro818.

Karmali, M.A., Gannon, V., and Sargeant, J.M. (2010). Verocytotoxin-producing Escherichia coli (VTEC). *Vet Microbiol* 140(3-4)**,** 360-370. doi: 10.1016/j.vetmic.2009.04.011.

Karmali, M.A., Mascarenhas, M., Shen, S., Ziebell, K., Johnson, S., Reid-Smith, R., *et al.* (2003). Association of genomic O island 122 of Escherichia coli EDL 933 with verocytotoxin-producing Escherichia coli seropathotypes that are linked to epidemic and/or serious disease. *J Clin Microbiol* 41(11)**,** 4930-4940.

Kimmitt, P.T., Harwood, C.R., and Barer, M.R. (2000). Toxin gene expression by shiga toxin-producing Escherichia coli: the role of antibiotics and the bacterial SOS response. *Emerg Infect Dis* 6(5)**,**

458-465. doi: 10.3201/eid0605.000503.

Konadu, E., Donohue-Rolfe, A., Calderwood, S.B., Pozsgay, V., Shiloach, J., Robbins, J.B., *et al.* (1999). Syntheses and immunologic properties of Escherichia coli O157 O-specific polysaccharide and Shiga toxin 1 B subunit conjugates in mice. *Infect Immun* 67(11)**,** 6191-6193.

Konadu, E., Robbins, J.B., Shiloach, J., Bryla, D.A., and Szu, S.C. (1994). Preparation, characterization, and immunological properties in mice of Escherichia coli O157 O-specific polysaccharide-protein conjugate vaccines. *Infect Immun* 62(11)**,** 5048-5054.

Konadu, E.Y., Parke, J.C., Jr., Tran, H.T., Bryla, D.A., Robbins, J.B., and Szu, S.C. (1998). Investigational vaccine for Escherichia coli O157: phase 1 study of O157 O-specific polysaccharide-Pseudomonas aeruginosa recombinant exoprotein A conjugates in adults. *J Infect Dis* 177(2)**,** 383-387.

Kotloff, K.L., Nataro, J.P., Blackwelder, W.C., Nasrin, D., Farag, T.H., Panchalingam, S., *et al.* (2013). Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet* 382(9888)**,** 209-222. doi: 10.1016/S0140-6736(13)60844-2.

Kovacs, S.D., Mullholland, K., Bosch, J., Campbell, H., Forouzanfar, M.H., Khalil, I., *et al.* (2015). Deconstructing the differences: a comparison of GBD 2010 and CHERG's approach to estimating the mortality burden of diarrhea, pneumonia, and their etiologies. *BMC Infect Dis* 15**,** 16. doi: 10.1186/s12879-014-0728-4.

Leitner, D.R., Lichtenegger, S., Temel, P., Zingl, F.G., Ratzberger, D., Roier, S., *et al.* (2015). A combined vaccine approach against Vibrio cholerae and ETEC based on outer membrane vesicles. *Front Microbiol* 6**,** 823. doi: 10.3389/fmicb.2015.00823.

Lewis, S.B., Cook, V., Tighe, R., and Schuller, S. (2015). Enterohemorrhagic Escherichia coli colonization of human colonic epithelium in vitro and ex vivo. *Infect Immun* 83(3)**,** 942-949. doi: 10.1128/IAI.02928-14.

Liu, J., Sun, Y., Feng, S., Zhu, L., Guo, X., and Qi, C. (2009). Towards an attenuated enterohemorrhagic Escherichia coli O157:H7 vaccine characterized by a deleted ler gene and containing apathogenic Shiga toxins. *Vaccine* 27(43)**,** 5929-5935. doi: 10.1016/j.vaccine.2009.07.097.

Liu, M., Ruan, X., Zhang, C., Lawson, S.R., Knudsen, D.E., Nataro, J.P., *et al.* (2011). Heat-labile- and heat-stable-toxoid fusions (LTR(1)(9)(2)G-STaP(1)(3)F) of human enterotoxigenic Escherichia coli elicit neutralizing antitoxin antibodies. *Infect Immun* 79(10)**,** 4002-4009. doi: 10.1128/IAI.00165-11.

Lundgren, A., Bourgeois, L., Carlin, N., Clements, J., Gustafsson, B., Hartford, M., *et al.* (2014). Safety and immunogenicity of an improved oral inactivated multivalent enterotoxigenic Escherichia coli (ETEC) vaccine administered alone and together with dmLT adjuvant in a double-blind, randomized, placebo-controlled Phase I study. *Vaccine* 32(52)**,** 7077-7084. doi:

10.1016/j.vaccine.2014.10.069.

MacLennan, C.A., and Saul, A. (2014). Vaccines against poverty. *Proc Natl Acad Sci U S A* 111(34), 12307-12312. doi: 10.1073/pnas.1400473111.

Mayr, U.B., Haller, C., Haidinger, W., Atrasheuskaya, A., Bukin, E., Lubitz, W., *et al.* (2005). Bacterial ghosts as an oral vaccine: a single dose of Escherichia coli O157:H7 bacterial ghosts protects mice against lethal challenge. *Infect Immun* 73(8), 4810-4817. doi: 10.1128/IAI.73.8.4810-4817.2005.

Mayr, U.B., Kudela, P., Atrasheuskaya, A., Bukin, E., Ignatyev, G., and Lubitz, W. (2012). Rectal single dose immunization of mice with Escherichia coli O157:H7 bacterial ghosts induces efficient humoral and cellular immune responses and protects against the lethal heterologous challenge. *Microb Biotechnol* 5(2), 283-294. doi: 10.1111/j.1751-7915.2011.00316.x.

McKenzie, R., Bourgeois, A.L., Engstrom, F., Hall, E., Chang, H.S., Gomes, J.G., *et al.* (2006). Comparative safety and immunogenicity of two attenuated enterotoxigenic Escherichia coli vaccine strains in healthy adults. *Infect Immun* 74(2), 994-1000. doi: 10.1128/IAI.74.2.994-1000.2006.

McKenzie, R., Bourgeois, A.L., Frech, S.A., Flyer, D.C., Bloom, A., Kazempour, K., *et al.* (2007). Transcutaneous immunization with the heat-labile toxin (LT) of enterotoxigenic Escherichia coli (ETEC): protective efficacy in a double-blind, placebo-controlled challenge study. *Vaccine* 25(18), 3684-3691. doi: 10.1016/j.vaccine.2007.01.043.

McKenzie, R., Darsley, M., Thomas, N., Randall, R., Carpenter, C., Forbes, E., *et al.* (2008). A double-blind, placebo-controlled trial to evaluate the efficacy of PTL-003, an attenuated enterotoxigenic E. coli (ETEC) vaccine strain, in protecting against challenge with virulent ETEC. *Vaccine* 26(36), 4731-4739. doi: 10.1016/j.vaccine.2008.06.064.

McWilliams, B.D., and Torres, A.G. (2014). Enterohemorrhagic Escherichia coli Adhesins. *Microbiol Spectr* 2(3). doi: 10.1128/microbiolspec.EHEC-0003-2013.

Mejias, M.P., Hiriart, Y., Lauche, C., Fernandez-Brando, R.J., Pardo, R., Bruballa, A., *et al.* (2016). Development of camelid single chain antibodies against Shiga toxin type 2 (Stx2) with therapeutic potential against Hemolytic Uremic Syndrome (HUS). *Sci Rep* 6, 24913. doi: 10.1038/srep24913.

Messens, W., Bolton, D., Frankel, G., Liebana, E., Mc, L.J., Morabito, S., *et al.* (2015). Defining pathogenic verocytotoxin-producing Escherichia coli (VTEC) from cases of human infection in the European Union, 2007-2010. *Epidemiol Infect* 143(8), 1652-1661. doi: 10.1017/S095026881400137X.

Monteiro, R., Ageorges, V., Rojas-Lopez, M., Schmidt, H., Weiss, A., Bertin, Y., *et al.* (2016). A secretome view of colonisation factors in Shiga toxin-encoding Escherichia coli (STEC): from enterohaemorrhagic E. coli (EHEC) to related enteropathotypes. *FEMS Microbiol Lett* 363(16). doi: 10.1093/femsle/fnw179.

Montero, D., Orellana, P., Gutierrez, D., Araya, D., Salazar, J.C., Prado, V., *et al.* (2014). Immunoproteomic analysis to identify Shiga toxin-producing Escherichia coli outer membrane proteins expressed during human infection. *Infect Immun* 82(11), 4767-4777. doi: 10.1128/IAI.02030-14.

Mora, M., Veggi, D., Santini, L., Pizza, M., and Rappuoli, R. (2003). Reverse vaccinology. *Drug Discov Today* 8(10), 459-464.

Moriel, D.G., Bertoldi, I., Spagnuolo, A., Marchi, S., Rosini, R., Nesta, B., *et al.* (2010). Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic Escherichia coli. *Proc Natl Acad Sci U S A* 107(20), 9072-9077. doi: 10.1073/pnas.0915077107.

Nataro, J.P., and Kaper, J.B. (1998). Diarrheagenic Escherichia coli. *Clin Microbiol Rev* 11(1), 142-201.

Nesta, B., Spraggon, G., Alteri, C., Moriel, D.G., Rosini, R., Veggi, D., *et al.* (2012). FdeC, a novel broadly conserved Escherichia coli adhesin eliciting protection against urinary tract infections. *MBio* 3(2). doi: 10.1128/mBio.00010-12.

Nesta, B., Valeri, M., Spagnuolo, A., Rosini, R., Mora, M., Donato, P., *et al.* (2014). SslE elicits functional antibodies that impair in vitro mucinase activity and in vivo colonization by both intestinal and extraintestinal Escherichia coli strains. *PLoS Pathog* 10(5), e1004124. doi: 10.1371/journal.ppat.1004124.

Noguera-Obenza, M., Ochoa, T.J., Gomez, H.F., Guerrero, M.L., Herrera-Insua, I., Morrow, A.L., *et al.* (2003). Human milk secretory antibodies against attaching and effacing Escherichia coli antigens. *Emerg Infect Dis* 9(5), 545-551.

Norton, E.B., Lawson, L.B., Freytag, L.C., and Clements, J.D. (2011). Characterization of a mutant Escherichia coli heat-labile toxin, LT(R192G/L211A), as a safe and effective oral adjuvant. *Clin Vaccine Immunol* 18(4), 546-551. doi: 10.1128/CVI.00538-10.

Ochoa, T.J., Barletta, F., Contreras, C., and Mercado, E. (2008). New insights into the epidemiology of enteropathogenic Escherichia coli infection. *Trans R Soc Trop Med Hyg* 102(9), 852-856. doi: 10.1016/j.trstmh.2008.03.017.

Ochoa, T.J., and Contreras, C.A. (2011). Enteropathogenic escherichia coli infection in children. *Curr Opin Infect Dis* 24(5), 478-483. doi: 10.1097/QCO.0b013e32834a8b8b.

Oliveira, A.F., Cardoso, S.A., Almeida, F.B., de Oliveira, L.L., Pitondo-Silva, A., Soares, S.G., *et al.* (2012). Oral immunization with attenuated Salmonella vaccine expressing Escherichia coli O157:H7 intimin gamma triggers both systemic and mucosal humoral immunity in mice. *Microbiol Immunol* 56(8), 513-522. doi: 10.1111/j.1348-0421.2012.00477.x.

Pacheco, A.R., and Sperandio, V. (2012). Shiga toxin in enterohemorrhagic E.coli: regulation and novel anti-virulence strategies. *Front Cell Infect Microbiol* 2, 81. doi: 10.3389/fcimb.2012.00081.

Parissi-Crivelli, A., Parissi-Crivelli, J.M., and Giron, J.A. (2000). Recognition of enteropathogenic

Escherichia coli virulence determinants by human colostrum and serum antibodies. *J Clin Microbiol* 38(7)**,** 2696-2700.

Pianciola, L., D'Astek, B.A., Mazzeo, M., Chinen, I., Masana, M., and Rivas, M. (2016). Genetic features of human and bovine Escherichia coli O157:H7 strains isolated in Argentina. *Int J Med Microbiol*. doi: 10.1016/j.ijmm.2016.02.005.

Pires, S.M., Fischer-Walker, C.L., Lanata, C.F., Devleesschauwer, B., Hall, A.J., Kirk, M.D., *et al.* (2015). Aetiology-Specific Estimates of the Global and Regional Incidence and Mortality of Diarrhoeal Diseases Commonly Transmitted through Food. *PLoS One* 10(12)**,** e0142927. doi: 10.1371/journal.pone.0142927.

Pizza, M., Scarlato, V., Masignani, V., Giuliani, M.M., Arico, B., Comanducci, M., *et al.* (2000). Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 287(5459)**,** 1816-1820.

Plotkin, S. (2014). History of vaccination. *Proc Natl Acad Sci U S A* 111(34)**,** 12283-12287. doi: 10.1073/pnas.1400472111.

Porter, C.K., Riddle, M.S., Alcala, A.N., Sack, D.A., Harro, C., Chakraborty, S., *et al.* (2016). An Evidenced-Based Scale of Disease Severity following Human Challenge with Enteroxigenic Escherichia coli. *PLoS One* 11(3)**,** e0149358. doi: 10.1371/journal.pone.0149358.

Quintana Flores, V.M., Campos de Souza Fernandes, R.C., Sousa de Macedo, Z., and Medina-Acosta, E. (2002). Expression and purification of the recombinant enteropathogenic Escherichia coli vaccine candidates BfpA and EspB. *Protein Expr Purif* 25(1)**,** 16-22. doi: 10.1006/prep.2001.1604.

Riquelme-Neira, R., Rivera, A., Saez, D., Fernandez, P., Osorio, G., Del Canto, F., *et al.* (2015). Vaccination with DNA Encoding Truncated Enterohemorrhagic Escherichia coli (EHEC) Factor for Adherence-1 Gene (efa-1') Confers Protective Immunity to Mice Infected with E. coli O157:H7. *Front Cell Infect Microbiol* 5**,** 104. doi: 10.3389/fcimb.2015.00104.

Rivas, M., Chinen, I., and Guth, B.E.C. (2016). "Enterohemorrhagic (Shiga Toxin-Producing) Escherichia coli," in *Escherichia coli in the Americas,* ed. A.G. Torres. (Cham: Springer International Publishing), 97-123.

Roy, K., Bartels, S., Qadri, F., and Fleckenstein, J.M. (2010). Enterotoxigenic Escherichia coli elicits immune responses to multiple surface proteins. *Infect Immun* 78(7)**,** 3027-3035. doi: 10.1128/IAI.00264-10.

Roy, K., Hamilton, D.J., Munson, G.P., and Fleckenstein, J.M. (2011). Outer membrane vesicles induce immune responses to virulence proteins and protect against colonization by enterotoxigenic Escherichia coli. *Clin Vaccine Immunol* 18(11)**,** 1803-1808. doi: 10.1128/CVI.05217-11.

Scaletsky, I.C.A., and Fagundes-Neto, U. (2016). "Typical Enteropathogenic Escherichia coli," in *Escherichia coli in the Americas,* ed. A.G. Torres. (Cham: Springer International Publishing), 59-76.

Sjoling, A., von Mentzer, A., and Svennerholm, A.M. (2015). Implications of enterotoxigenic Escherichia coli genomics for vaccine development. *Expert Rev Vaccines* 14(4), 551-560. doi: 10.1586/14760584.2015.996553.

Stacy M. Crim, P.M.G., Robert Tauxe, Ellyn P. Marder, Debra Gilliss, Alicia B. Cronquist, , and Matthew Cartter, M.T.-D.A., David Blythe, Kirk Smith, Sarah Lathrop, Shelley Zansky, Paul R. Cieslak, John Dunn, Kristin G. Holt, Beverly Wolpert, Olga L. Henao. (2015). "Preliminary Incidence and Trends of Infection with Pathogens Transmitted Commonly Through Food — Foodborne Diseases Active Surveillance Network, 10 U.S. Sites, 2006–2014", in: *MMWR Morb Mortal Wkly Rep.* (ed.) M.a.M.W.R. (MMWR). (Atlanta: Center for Surveillance, Epidemiology, and Laboratory Services, Centers for Disease Control and Prevention (CDC)).

Tacket, C.O., Sztein, M.B., Losonsky, G., Abe, A., Finlay, B.B., McNamara, B.P., *et al.* (2000). Role of EspB in experimental human enteropathogenic Escherichia coli infection. *Infect Immun* 68(6), 3689-3695.

Tapia, D., Ross, B.N., Kalita, A., Kalita, M., Hatcher, C.L., Muruato, L.A., *et al.* (2016). From In silico Protein Epitope Density Prediction to Testing Escherichia coli O157:H7 Vaccine Candidates in a Murine Model of Colonization. *Front Cell Infect Microbiol* 6, 94. doi: 10.3389/fcimb.2016.00094.

Tarr, P.I., Gordon, C.A., and Chandler, W.L. (2005). Shiga-toxin-producing Escherichia coli and haemolytic uraemic syndrome. *Lancet* 365(9464), 1073-1086. doi: 10.1016/S0140-6736(05)71144-2.

Taylor, K.A., O'Connell, C.B., Luther, P.W., and Donnenberg, M.S. (1998). The EspB protein of enteropathogenic Escherichia coli is targeted to the cytoplasm of infected HeLa cells. *Infect Immun* 66(11), 5501-5507.

Tobias, J., Holmgren, J., Hellman, M., Nygren, E., Lebens, M., and Svennerholm, A.M. (2010). Over-expression of major colonization factors of enterotoxigenic Escherichia coli, alone or together, on non-toxigenic E. coli bacteria. *Vaccine* 28(43), 6977-6984. doi: 10.1016/j.vaccine.2010.08.047.

Tobias, J., Lebens, M., Bolin, I., Wiklund, G., and Svennerholm, A.M. (2008). Construction of non-toxic Escherichia coli and Vibrio cholerae strains expressing high and immunogenic levels of enterotoxigenic E. coli colonization factor I fimbriae. *Vaccine* 26(6), 743-752. doi: 10.1016/j.vaccine.2007.12.009.

Tobias, J., Svennerholm, A.M., Carlin, N.I., Lebens, M., and Holmgren, J. (2011). Construction of a non-toxigenic Escherichia coli oral vaccine strain expressing large amounts of CS6 and inducing strong intestinal and serum anti-CS6 antibody responses in mice. *Vaccine* 29(48), 8863-8869. doi: 10.1016/j.vaccine.2011.09.096.

Torres, A.G. (2017a). Escherichia coli diseases in Latin America-a 'One Health' multidisciplinary approach. *Pathog Dis* 75(2). doi: 10.1093/femspd/ftx012.

Torres, A.G. (2017b). Maternal immunity, a wayto confer protection against enteropathogenic Escherichia coli. *J Pediatr (Rio J).* doi: 10.1016/j.jped.2017.05.002.

Turner, A.K., Stephens, J.C., Beavis, J.C., Greenwood, J., Gewert, C., Randall, R., *et al.* (2011). Generation and characterization of a live attenuated enterotoxigenic Escherichia coli combination vaccine expressing six colonization factors and heat-labile toxin subunit B. *Clin Vaccine Immunol* 18(12)**,** 2128-2135. doi: 10.1128/CVI.05345-11.

Turner, A.K., Terry, T.D., Sack, D.A., Londono-Arcila, P., and Darsley, M.J. (2001). Construction and characterization of genetically defined aro omp mutants of enterotoxigenic Escherichia coli and preliminary studies of safety and immunogenicity in humans. *Infect Immun* 69(8)**,** 4969-4979. doi: 10.1128/IAI.69.8.4969-4979.2001.

Vidal, R.M., Chamorro, N.L., and Girón, J.A. (2016). "Enterotoxigenic Escherichia coli," in *Escherichia coli in the Americas,* ed. A.G. Torres. (Cham: Springer International Publishing), 1-26.

Wan, C.S., Zhou, Y., Yu, Y., Peng, L.J., Zhao, W., and Zheng, X.L. (2011). B-cell epitope KT-12 of enterohemorrhagic Escherichia coli O157:H7: a novel peptide vaccine candidate. *Microbiol Immunol* 55(4)**,** 247-253. doi: 10.1111/j.1348-0421.2011.00316.x.

Wen, S.X., Teel, L.D., Judge, N.A., and O'Brien, A.D. (2006). A plant-based oral vaccine to protect against systemic intoxication by Shiga toxin type 2. *Proc Natl Acad Sci U S A* 103(18)**,** 7082-7087. doi: 10.1073/pnas.0510843103.

Whitfield, C. (2006). Biosynthesis and assembly of capsular polysaccharides in Escherichia coli. *Annu Rev Biochem* 75**,** 39-68. doi: 10.1146/annurev.biochem.75.103004.142545.

Zhang, C., Iqbal, J., and Gomez-Duarte, O.G. (2016). Murine immunization with CS21 pili or LngA major subunit of enterotoxigenic Escherichia coli (ETEC) elicits systemic and mucosal immune responses and inhibits ETEC gut colonization. *Vet Microbiol.* doi: 10.1016/j.vetmic.2016.02.001.

Zhang, X.H., He, K.W., Zhang, S.X., Lu, W.C., Zhao, P.D., Luan, X.T., *et al.* (2011). Subcutaneous and intranasal immunization with Stx2B-Tir-Stx1B-Zot reduces colonization and shedding of Escherichia coli O157:H7 in mice. *Vaccine* 29(22)**,** 3923-3929. doi: 10.1016/j.vaccine.2011.02.007.

Table 1. Summary of the most prospective vaccine cadiadates against EHEC and ETEC.

EHEC			
Type	**Key component**	**Efficacy outcome**	**Reference**
Shiga toxin-based vaccines	Antibodies cαStx1B and cαStx2A	Good tolerance and safety in a human trial single dose study.	Bitzan *et al.*, 2009
	Two copies of anti-Stx2B VHH and one anti-sero albumin VHH	Decrease the toxicity in Stx2 lethal mouse model.	Mejias *et al.*, 2016
Attenuated bacteria-based vaccines	EHEC O157:H7 86-24 strain Δ*ler*Δ*stx2* expressing Stx1A Stx2A detoxified	Reduce EHEC O157:H7 wild type strain colonization.	Liu *et al.*, 2009
	Attenuated EPEC O126:H6	Reduce mortality in mice model and shown cross-reaction of produced EspB and intimin EPEC antibodies with EspB and intimin from EHEC.	Calderon Toledo *et al.*, 2011
	Attenuated *Salmonella* Typhimurium expressing recombinant EspA, Intimin and Stx2B	Higher titer of specific antibodies, and specific lymphocyte proliferation.	Oliveira *et al.*, 2012
	Attenuated *Salmonella* Typhimurium χ3987 (Δ*cya*,Δ*crp*,Δ*asd* and H683 Δ*aro*Δ*asd*) expressing γ-intimin variant	Induce systemic and humoral immunity by increase titers of IgG in serum and IgA in feces. Reduced the EHEC O157:H7 shedding post challenge.	Oliveira *et al.*, 2012
	Recombinant Bacillus Calmette-Guérin expressing Stx2B (rBCG-Stx2B)	Increase the levels of Stx2 IgG in mice. Higher survival rate (63%) in immunized mice after EHEC challenge.	Fujii *et al.*, 2012
Bacterial Ghost-based vaccines	Bacterial Ghosts of O157:H7 unable to cause infection	Antitoxicity effect on Vero cells culture. Stop the shedding of EHEC O157:H7wild type strain and rate survival of 93% in orally immunized mice and 100% in rectal immunized mice.	Mayr *et al.*, 2012
	Bacterial Ghosts of O157:H7 exposing Stx chimerical protein (Stx2Am-Stx1B)	High specific IgG and IgA antibodies titer to Stx1A and Stx2B. Survival rate of 52% in immunized mice.	Cai *et al.*, 2015
Protein-based vaccines	EspA-Stx1A fusion protein	High titters of specific IgG antibodies to EspA-Stx1A and 95% of mice survival after a challenge with crude toxin Stx2.	Cheng *et al.*, 2009
	Stx1B-Stx2-truncated intimin fusion protein	100% of survival rate of orally immunized mice challenged with EHEC O157:H7 88321 and anti-toxin and anti-adhesion	Gansheroff *et al.*, 1999

		effect.	
	Stx2Am-Stx1B, SAmB fusion protein	93% survival rate of orally immunized mice challenged with EHEC O157:H7 88321.	Gao *et al.*, 2011
Peptide-based vaccines	C terminal region of intimin	Reduce the bacterial attachment to Hep-2 cells and confers protection to immunized mice.	Wan *et al.*, 2011
	Peptide KT-12 (KASITEIKADKT) conjugated with KLH	High concentrations of IgG in subcutaneous immunized mice and high titer of IgA in intranasal immunized mice.	Zhang *et al.*, 2011
Plant based vaccines	*Nicotiana tabacum* (tobacco) NT-1 cell line expression inactivated Stx1A	High specific IgA anti Stx2 in fecal samples from orally immunized mice, conferring high protection against STEC strain B2F1 (75% survival rate).	Wen *et al.*, 2006
	Chimerical gene *espA-eae-tir*	Reduction of EHEC O157:H7 shedding, colonization and histological damage in subcutaneously or orally immunized mice.	Amani *et al.*, 2009
DNA-based vaccines	Stx2AΔAB DNA vaccine	Protection to immunized mice challenged with native Stx2 and toxin neutralization in Vero culture cells.	Bentancor *et al.*, 2009
	C-terminal domain of *EscC*	Reduction bacterial counts in feces, colon and caecum and increment of IgGs in sera and IgA in feces from intranasal immunized mice.	Garcia-Angulo *et al.*, 2014 Tapia *et al.*, 2016
	pVAX-efa1 (*efa-1'*)	High levels of specific mucosal IgA and reduction of EHEC colonization.	(Riquelme-Neira *et al.*, 2015
Polysaccharide-based vaccines	O-specific polysaccharide of EHEC O157:H7 conjugated with recombinant exotoxin A of *P. aeruginosa* (O157-rEPA)	High levels of IgG against LPS in vaccinated children with non-collateral reactions to the vaccine.	Konadu *et al.*, 1994 Ahmed *et al.*, 2006
Adjuvant enhanced vaccines	EspB and/or C-terminal of γ-Intimin proteins + MALP-2 adjuvant	Higher titers of IgA at broncheo-alveolar level.	Cataldi *et al.*, 2008
	Chimerical protein Tir-Stx1B-Stx2B + Zot adjuvant	Protection against EHEC. High IgA and IgG response and reduction of bacterial shedding in faces post EHEC wild-type challenge in subcutaneous immunized mice.	Zhang *et al.*, 2011
ETEC			
Toxin-based vaccines	LT toxin using skin patch	Increase amount of anti-LT IgG and IgA in 97-100% of human volunteers.	McKenzie *et al.*, 2007
	STaP13F-LTR192G toxoid	Development of IgG specific	Liu *et al.*,

	fusion protein	antibodies for LT and STa proteins in serum and feces and only IgA in feces in immunized mice.	2011
Autotransporters-based vaccines	Recombinant Ag43 and pAT autotransporters	Increase of fecal IgA and relative protection against intestinal colonization in immunized mice.	Harris *et al.*, 2011
Adhesins-base vaccines	Recombinant ETEC two-partner secretion protein A (EtpA)	Inhibition of ETEC colonization in immunized mice.	Roy *et al.*, 2009
	CS21/LngA formulated with cholera toxin	Increase specific IgG antibodies in serum and IgA antibodies in fecal and intestinal lavages and stop the bacteria shedding in immunized mice.	Zhang *et al.*, 2016
Attenuated bacteria-based vaccines	ETEC E1392/75-2A Δ*aroC*Δ*ompR* and ETEC E1392/75-2A Δ*aroC*Δ*ompR*Δ*ompC*	IgA and IgG response and specific antibodies against CS1 and CS3.	Turner *et al.*, 2001 McKenzie *et al.*, 2006
	Etvax (attenuated bacteria expressing CS6 in K12 and CFA/I, CS3, CS5 in ETEC O78 toxin-negative) + LCTBA hybrid protein + double mutant LT	High titers of fecal, jejunal and serum IgA and IgG in orally immunized human.	Norton *et al.*, 2011
	Vibrio cholerae strains expressing CFA/I operon (cfaA-B-C-E)	Increase IgA and IgG titers in serum of immunized mice.	Tobias *et al.*, 2008
	ETEC strains expressinCFA/I (ACAM2010 str.), CS2 and CS3 (ACAM2007 str.) and CS1, CS2, and CS3 (ACAM2017 str.)	Increase the IgA against CFAs level in human volunteers orally immunized.	Daley *et al.*, 2007
	Non-toxigenic *E. coli* expressing CS2, CS4, CS5, or CS6 and CFA/I	Induce IgG+IgM antibodies against CS6 in sera and IgA in fecal samples of immunized mice.	Tobias *et al.*, 2010
	ACE527 ETEC complex (ACAM2022 (O141:H5, expressing CS5 and CS6), ACAM2025 (O39:H12, expressing CFA/I) and ACAM2027 (O71:H-, expressing CS2, CS3, and CS1)	Help to short the duration of the diarrhea and reduce the shedding of the wild-type strain, conferring a protection level ranging between 33% to 98% in human clinical trials.	Harro *et al.*, 2011 Harro *et al.*, 2011b
OMVs-based vaccines	ETEC OMVs Δ*msbB*Δ*eltA*	Detoxified OMV yielded higher titers of IgG1, IgM, and IgA and reduce wild type colonization in immunized mice.	Leitner *et al.*, 2015
	Vibrio cholerae OMV Δ*msbB*Δ*ctxAB*Δ*flaA* expressing ETEC FliC and CFA/I	OMV yielded higher titers of IgG1, IgM, and IgA and reduce wild type colonization in immunized mice.	Leitner *et al.*, 2015

Résumé long français

Cette thèse de doctorat s'est déroulée dans le cadre d'un projet européen FP7 (7th Framework Program) MSCA (Marie Sklodowska-Curie action) ITN (Initiale Training Network) EID (European Industrial Doctorates) appelé DIS*Co* (a multidisciplinary Doctoral Industrial School on novel preventive strategies against *Escherichia coli* infections) coordonné par Mariagrazia Pizza et co-coordonné par Mickaël Desvaux. Ainsi, ce doctorat s'est déroulé pour moitié en Italie au centre de recherche GSK (GlaxoSmithKline) sur le site de Sienne sous la supervision de Roberto Rosini et la direction de Fabio Polticelli de Universita degli Studi Roma Tre. L'autre moitié de la thèse s'est déroulée en France à l'INRA, centre Auvergne-Rhône-Alpes sur le site de Theix sous la direction de Mickaël Desvaux et Grégory Jubelin comme co-encadrant. Cette thèse de doctorat participe au développement de nouvelles stratégies préventives aux infections aux *E. coli* pathogènes intestinaux (InPEC), en particulier *E. coli* entérohémorragiques (EHEC), par une stratégie vaccinale. Dans ce contexte, une approche de vaccinologie inverse a été mise en œuvre pour identifier de nouveaux antigènes candidats qui ont ensuite été délivrés par la technologie GMMA (Generalized Modules for Membrane Antigens). Par ailleurs, un domaine épitope potentiel chez les autotransporteurs, *i.e.* l'autochaperon, a été caractérisé par analyse des séquences protéiques et modélisation structurale.

INTRODUCTION

Escherichia coli est l'un des microorganismes les plus communs du microbiote intestinal. La colonisation bactérienne par *E. coli* commence depuis la naissance et développe une relation symbiotique (Nataro et Kaper, 1998, Kaper et al., 2004). Cependant, certaines souches d'*E. coli* ont évolué en bactéries pathogènes. Selon le site d'infection, ces *E. coli* pathogènes sont classés parmi les *E. coli* pathogènes intestinaux (InPEC) ou parmi les *E. coli* pathogènes extra-intestinaux (ExPEC). Tandis que les ExPEC comprennent les *E. coli* uropathogènes (UPEC) et les *E. coli* responsables de méningites néonatales (NMEC) (Nataro et Kaper, 1998b, Kaper et al., 2004), les InPECs sont subdivisés en six pathotypes : les *E. coli* entérotoxigènes (ETEC), les *E. coli* entéropathogènes (EPEC), les *E. coli* entérohémorragiques (EHEC), les *E. coli* entéroaggrégatifs (EAEC), les *E. coli* entéroinvasifs (EIEC), et les *E. coli* diffusifs adhérents (DAEC). Cette classification est basée sur des différences de caractéristiques génétiques et phénotypiques des souches, du tableau clinique de l'infection qu'elles provoquent, du

mécanisme de pathogénèse et de leur interaction avec l'entérocytes (Figure 1). De par l'expression de Shiga-toxines (Stx), les EHEC appartiennent au plus large groupe des *E. coli* codants/producteurs des Stx (STEC).

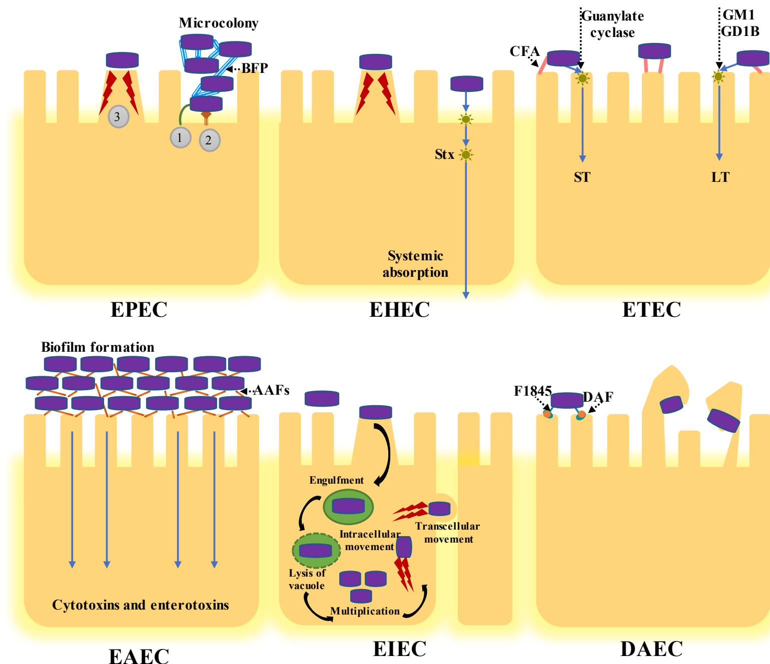


Figure 1 : **Représentation schématique des six pathotypes InPEC, selon leurs principaux facteurs de virulence et leur interaction avec les entérocytes** (Figure basée sur Kaper et al., 2004). Les EPEC colonisent les entérocytes de l'intestin grêle, détruisent l'architecture microvillaire et induisent des lésions d'attachement et d'effacement (A/E). Le remodelage cytosquelettique se produit au cours de la réponse inflammatoire et de l'apparition de diarrhée. 1 : Adhésion initiale, 2 : Translocation des protéines par le système de sécrétion de Type III, 3 : Formation d'une structure en piédestal. Les EHEC colonisent le côlon et induisent également des lésions A/E avec la caractéristique principale d'exprimer des Shiga-toxines (Stx). Les ETEC adhèrent aux entérocytes de l'intestin grêle provoquant une diarrhée aqueuse due à la sécrétion d'entérotoxines thermolabiles (LT) et/ou thermostables (ST). Les EAEC adhèrent à l'épithélium de l'intestin grêle et forme un biofilm épais avec l'expression de cytotoxines et d'entérotoxines sécrétées. Les EIEC pénètrent dans les cellules épithéliales du côlon, lysent le phagosome et se déplacent à l'intérieur des cellules infectées en restructurant les microfilaments d'actine. Les EIEC peuvent se déplacer latéralement à travers l'épithélium par propagation directe de cellule à cellule ou peuvent sortir et réintégrer la membrane plasmique baso-latérale. Les DAEC induisent une transduction de signal caractéristique dans les entérocytes de l'intestin grêle et le développement de longues projections cellulaires s'enroulant autour de la bactérie. AAF, fimbriae adhérent agrégative ; BFP, pilus formant des faisceaux ; CFA, facteur de colonisation antigénique ; DAF, facteur d'accélération de la décroissance ; ST, entérotoxine thermostable ; LT, entérotoxine thermolabile ; EAST1, entérotoxines thermostables d'*E. coli* entéroaggrégatifs ; ShET1, entérotoxine 1 de *Shigella*.

Les *E. coli* entérohémorragiques (EHEC) sont des pathogènes importants associés à plusieurs

épidémies diarrhéiques, principalement dues à la consommation de nourritures ou de boissons contaminées par ces bactéries, notamment de la viande hachée, des végétaux, jus de pomme ou encore des produits laitiers non pasteurisés. Les bovins sont considérés comme le principal réservoir naturel des EHEC où le portage est asymptomatique. En tant qu'anthropozoonose, les infections à EHEC peuvent provenir d'une contamination fécale directe (viande, lait) ou indirecte (légumes, eau) des produits alimentaires (Rivas et al., 2016). Les populations à risque comprennent les enfants et les personnes âgées. Aux États-Unis, en 2014, l'incidence pour 100 000 habitants était de 690 cas pour les EHEC O157 et de 445 pour les EHEC non O157. En ce qui concerne les EHEC non-O157:H7, leur incidence est en augmentation (Stacy M. Crim et Matthew Cartter, 2015). L'une des complications majeures suite aux infections aux EHEC est le syndrome hémolytique et urémique (SHU). Ce syndrome résulte principalement de l'expression et de l'action de Shiga-toxines (Stx), il est caractérisé par une triade d'anémie hémolytique microangiopathique, de thrombocytopénie et d'insuffisance rénale aiguë, avec le risque de développement de SHU (syndrome hémolytique et urémique) dans 10 % des cas chez les enfants de moins de 10 ans. Les principales épidémies recensées se sont produites dans les pays développés, comme les Etats-Unis d'Amérique, le Canada, l'Australie, le Japon, la France et d'autres en Europe. Néanmoins, les pays d'Amérique latine ont aussi connu des épisodes épidémiques où les EHEC posent ainsi un problème de santé publique, notamment en Argentine, au Brésil et au Chili (Nataro et Kaper, 1998b, Kaper et al., 2004, Caprioli et al., 2005, Mejias et al., 2016).

Entre EHEC O157 et non-O157, les manifestations cliniques sont très similaires et peuvent commencer par des diarrhées aqueuses légères avant d'évoluer en diarrhées sanglantes appelées colites hémorragiques. La période d'incubation avant l'apparition des premiers symptômes est d'environ 3 jours pour les EHEC O157:H7 (Bell et al., 1994). En plus des épisodes diarrhéiques, les principaux symptômes pouvant se manifester au début de l'infection sont de la fièvre, des crampes abdominales et/ou des vomissements (Tarr et al., 2005). Suite à l'apparition des colites hémorragiques, les patients sont généralement hospitalisés et peuvent guérir de l'infection après une semaine (Tarr et al., 2005). Cependant, il n'existe à ce jour aucun moyen de prévenir l'infection ni le développement vers des complications comme le SHU. Les traitements restent symptomatiques pour les infections à EHEC avec des perfusions intraveineuses d'électrolytes, le contrôle de la fonction rénale et des plaquettes, et où l'utilisation de médicaments antidiarrhéiques, antidouleurs et/ou antibiotiques est déconseillée (Holtz et al., 2009).
Comme pour tous les InPEC, les EHEC sont définis par un ensemble de preuves basées sur des

manifestations cliniques, notamment des symptômes, ainsi que des caractéristiques histologiques et moléculaires (Nataro et Kaper, 1998a, Kaper et al., 2004). Selon leur incidence, leur implication dans des épidémies ou leur association avec des formes plus ou moins sévères de la maladie, les STEC peuvent être classés en 4 classes, à savoir (i) la classe A correspondant au sérogroupe O157 avec une virulence élevée, (ii) la classe B comprenant les six sérotypes non O157:H7, *i.e.* O26:H11, O45:H2, O103:H2, O111:H8, O121:H19 et O145:H28, avec une virulence modérée, (iii) la classe C pour les sérotypes principalement impliqués dans certains cas sporadiques avec une virulence modérée, et (iv) classe D pour les sérotypes jamais rapportés dans l'infection humaine et donc avirulent (Karmali et al., 2003). Cette classification met aussi en exergue la différence entre EHEC et STEC, puisque les EHEC étant isolés de cas cliniques sont nécessairement pathogènes alors que les STEC ne le sont pas forcément ; tandis que les EHEC répondent à des critères phénotypiques, les STEC ne font références qu'à une caractéristique génotypique, c'est-à-dire la présence de gènes codant pour des Stx.

L'adhésion et la persistance de la bactérie est une caractéristique clé dans la pathogenèse bactérienne. Les EHEC O157:H7 colonisent principalement le côlon humain et portent un îlot de pathogénicité (PAI) connu sous le nom de locus d'effacement des entérocytes (LEE), aussi présent chez les EPEC où il est responsable des lésions A/E. Cependant, ces lésions ne sont jamais observées sur des biopsies de patients infectés par des EHEC mais ont seulement pu être observées lors d'essais *in vitro*. Par ailleurs, les souches d'EHEC non-O157:H7 ne portent pas nécessairement le LEE (McWilliams et Torres, 2014, Monteiro et al., 2016).

Les Stx correspondent aux facteurs de virulence caractéristiques des souches d'EHEC impliqués dans la pathogenèse et le développement du SHU. Les Stx sont codées à partir d'un prophage intégré dans le chromosome bactérien. Les Stx appartiennent à la famille AB et sont classées en deux types principaux, les Stx1 et Stx2. Les Stx1 se déclinent encore en 3 sous-types (a, c et d) et les Stx2 en 7 sous-types (de a à g) (Scheutz et al., 2012). Une souche d'EHEC peut coder pour un ou deux types de Stx, et certaines souches peuvent combiner différents types et sous-types de Stx. Néanmoins, Stx2 s'avère être la toxine associée aux formes les plus sévères d'infection aux EHEC (Boerlin et al., 1999). Globalement, les Stx provoquent des lésions cellulaires après leur internalisation en interagissant avec le récepteur Gb3 dans les cellules de Paneth et rénales provoquant la dépurination (adénine) de la sous-unité 28S du ribosome (ARNr 28S), qui entraîne alors l'arrêt de la synthèse des protéines. Ces toxines peuvent activer la cascade de coagulation, ainsi que l'hémolyse intravasculaire et ischémique (Nataro et Kaper,

1998b, Donnenberg et Whittam, 2001, Thorpe, 2004). La présence de la toxine dans la circulation sanguine entraînerait une aggravation de la maladie (Pacheco et Sperandio, 2012). Par ailleurs, il a été montré que l'utilisation d'antibiotiques entraînait la production de Stx (Zhang et al., 2000). Notamment, des concentrations sub-inhibitrices de fluoroquinolones ou de triméthoprime peuvent activer le système SOS induisant le cycle lytique du prophage conduisant à la lyse bactérienne et à la libération des Stx (Kimmitt et al., 2000, McGannon et al., 2010). Pour ces raisons, les traitements antibiotiques lors d'infections à EHEC ne sont pas recommandés et même proscrits.

Le plasmide pO157 est un méga-plasmide de 92 kb très commun parmi les souches de STEC/EHEC O157:H7. Il s'agit d'un plasmide de type F qui porte des gènes codant pour des facteurs de virulence importants, parmi lesquels (i) StcE, une métalloprotéase à zinc sécrétée par le système de sécrétion de type II (T2SS) lui–même codé par les gènes *etp*, (ii) l'hémolysine HlyA, (iii) la catalase peroxydase KatP, (iv) la sérine protéase EspP impliquée dans le clivage du facteur de coagulation V, ou encore (v) l'inhibiteur des stéroïdes C1 du complément (C1-INH) aussi impliqué dans la colonisation et la formation de lésions tissulaires par son activité mucinase (Burland et al., 1998; Donnenberg et Whittam, 2001; Torres, 2008).

L'adhésion d'*E. coli* O157:H7 et d'autres sérotypes EHEC aux cellules épithéliales de l'intestin est l'une des étapes cruciales de la pathogenèse. Le facteur d'adhésion le mieux étudié est le locus d'entérocyte et d'effacement (LEE). Cet îlot de pathogénicité de 35 Kb organisé en 5 opérons (du LEE-1 au LEE-5) confère le phénotype de lésion A/E essentiellement observé dans des modèles cellulaires *in vitro* ou des explants pour les EHEC/STEC (Nataro et Kaper, 1998). Ce phénotype est lié à la reconnaissance étroite entre les protéines Tir et intimin, codées par les gènes *tir* et *eae* respectivement. Néanmoins, ces protéines sont nécessaires mais pas suffisantes pour produire des lésions A/E typiques, ce qui suggère l'implication d'autres facteurs d'adhésion pour l'interaction initiale avec l'entérocyte (Torres et al., 2002; et al., 2005). Les séquençages complets de génomes de souches d'EHEC O157:H7 indiquent l'existence d'au moins une douzaine d'adhésines, dont (i) EspA, formant un filament sécrété et assemblé par un T3SS codé par le LEE, notamment régulé par ToxB (Tatsuno, I. et al, 2001, Stevens M.P. et al 2004), (ii) Iha (IrgA homologue adhesin) (Stevens, 2005) ou encore (iii) Cah (calcium binding antigen 43 homologue) (Torres et al., 2005). Concernant les pili (ou fimbriae), au moins seize opérons ont pu être identifiés (Perna et al., 2001) et certains caractérisés dont les (i) fimbriae de type 1 (McWilliams and Torres, 2014), (ii) curli (McWilliams et al, 2014), (iii) F9 fimbriae

(McWilliams and Torres, 2014), (iv) Ecp (*E. coli* common pili) (Rendon et al., 2007,), (v) pili de type 4 HCP (haemorrhagic *E. coli* pili) (Xicohtencatl-Cortes et al., 2007 and 2009), (vi) ELF (*E. coli* laminin-binding fimbriae) (Samadder et al., 2009), ou encore (vii) Lpf1 et Lpf2 (long polar fimbriae) (Torres et al., 2002, 2004). Chez les EHEC LEE-négatif, les adhésines les plus souvent retrouvées incluent (i) Lpf2 (LpfO113) (Doughty et al., 2002), (ii) Saa (STEC autoagglutinating adhesin) (Paton et al., 2001), (iii) Sab (STEC autotransporter contributing to biofilm formation) (Herold et al., 2009) ou encore Eib (*E. coli* immunoglobulin-binding protein) (Lu et al., 2006, Merkel et al., 2010).

Les facteurs de colonisation chez les *Escherichia coli* codant des Shiga toxines (STEC) : Depuis les *E. coli* entérohémorragiques (EHEC) jusqu'aux entéropathotypes associés (A secretome view of colonisation factors in Shiga toxin-encoding *Escherichia coli* (STEC): from enterohaemorrhagic *E. coli* (EHEC) to related enteropathotypes)
Article de revue (accepté dans FEMS Microbiology Letters)

Ricardo MONTEIRO [1,2], Valentin AGEORGES [1], Maricarmen ROJAS-LOPEZ [1,2], Herbert SCHMIDT [3], Agnes WEISS [3], Yolande BERTIN [1], Evelyne FORANO [1], Grégory JUBELIN [1], Ian HENDERSON [5], Valérie LIVRELLI [4], Alain P. GOBERT [1], Roberto ROSINI [2], Marco SORIANI [2] and Mickaël DESVAUX [1,*]

[1] INRA, UR454 Microbiologie, F-63122 Saint-Genès Champanelle, France, [2] GlaxoSmithKline Vaccines, Via Fiorentina 1, 53100, Siena, Italy, [3] Department of Food Microbiology and Hygiene, Institute of Food Science and Biotechnology, University of Hohenheim, Garbenstrasse 28, 70599 Stuttgart, Germany, [4] Centre de Recherche en Nutrition Humaine Auvergne, UMR INSERM UdA U1071, USC-INRA 2018, Clermont Université - Université d'Auvergne, Faculté de Pharmacie, CHU Clermont-Ferrand, Service Bactériologie Mycologie Parasitologie, Clermont-Ferrand, France, [5] Institute of Microbiology and Infection, University of Birmingham, Birmingham, B15 2TT, United Kingdom. * Auteur correspondant.

La définition du groupe des *Escherichia coli* produisant des Shiga toxines (STEC), qui ne sont pas nécessairement pathogènes, est en fait essentiellement basée sur la présence de gènes codant pour des Shiga toxines (Stx). Parmi les *E. coli* pathogènes intestinaux (InPEC), les *E. coli* entérohémorragiques (EHEC) constituent le sous-groupe majeur des STEC virulents. Les EHEC causent des maladies humaines graves telles que la colite hémorragique et le syndrome

hémolytique et urémique (SHU). Alors que les EHEC ont évolué à partir d'*E. coli* entéropathogènes (EPEC), des hybrides avec les *E. coli* entéroagrégatifs (EAEC) sont récemment apparus. Il est à noter que certains *E. coli* entéroinvasifs (EIEC) appartiennent également au groupe STEC. Alors que le LEE (locus de l'effacement des entérocytes) est un déterminant moléculaire important dans la pathogénicité, ni tous les EHEC ni tous les STEC ne contiennent le LEE, suggérant qu'ils possèdent des facteurs de virulence et de colonisation supplémentaires. Actuellement, neuf systèmes de sécrétion protéique ont été décrits chez des bactéries lipopolysaccharidiques didermes (LPS) (archétype Gram négatif) et peuvent être impliqués dans la sécrétion d'effecteurs extracellulaires, de protéines de surface ou d'assemblage d'organelles, comme le flagelle ou pili. Certains composants du sécrétome des STEC et des entéropathotypes apparentés sont impliqués dans la colonisation des surfaces biotiques et abiotiques (Figure 2). La richesse des mécanismes impliqués dans le trafic protéique, la combinaison des facteurs de colonisation ainsi que la modulation de leur expression soulignent l'importance du sécrétome dans l'écophysiologie des STEC. Il s'avère pertinent de considérer la diversité des facteurs de colonisation dans les différents sous-groupes de STEC et des entéropathotypes d'*E. coli* apparentés par rapport aux voies et systèmes de sécrétion des protéines chez les bactéries lipopolysaccharidiques didermes (bactéries à Gram négatif).

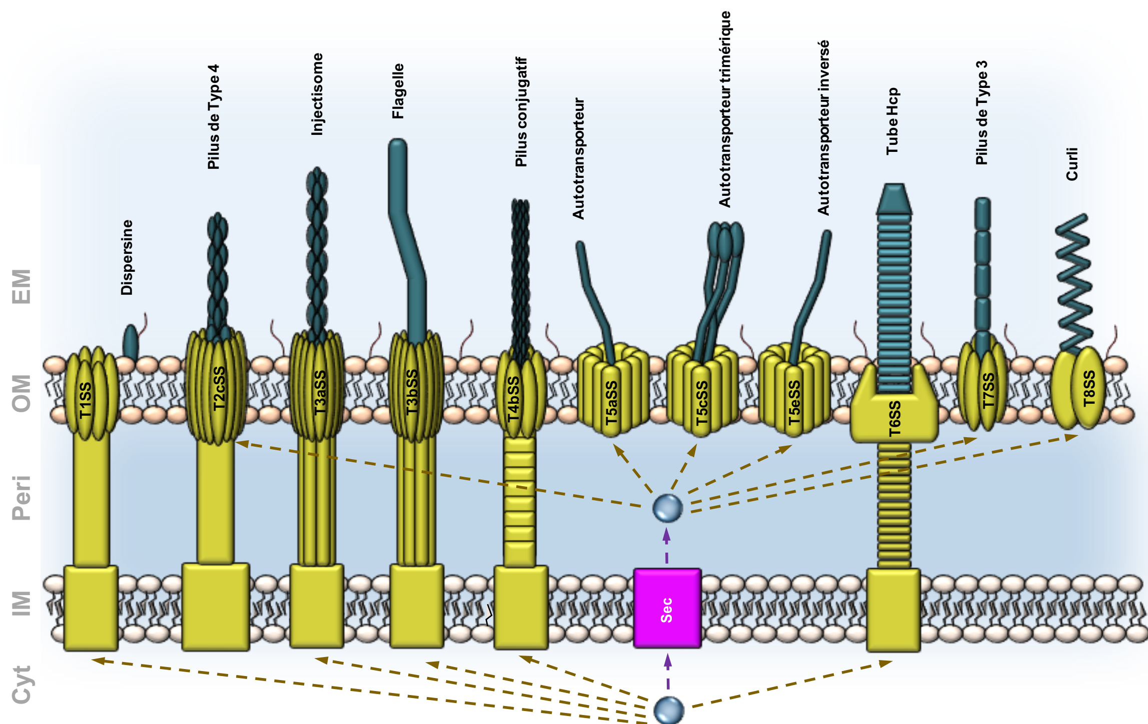


Figure 2 : **Vue d'ensemble des composants du sécrétome des STEC associés à la colonisation de surface**. Les systèmes d'exportation des protéines sont colorés en violet, les systèmes de sécrétion de protéines sont colorés en vert. Les effecteurs associés à chaque système de sécrétion et jouant le rôle de facteurs de colonisation sont colorés en bleu. T1SS : Système de sécrétion protéique de Type I ou système hétérotrimérique ABC-MFP-TolC. T2SS : Système de sécrétion protéique de Type II ou voie sécréton-dépendante (SDP), en particulier le système T4P (pili de type 4) correspondant au T2SS sous-type c (T2cSS). T3SS : Système de sécrétion protéique de Type III, incluant l'injectisome (T3aSS) et le flagelle (T3bSS). T4SS : Système de sécrétion protéique de Type IV, en particulier le système de conjugaison Tra (T4bSS). T5SS : Système de sécrétion protéique de Type V, comprenant notamment les autotransporteurs (T5aSS), les autotransporteurs trimériques (T5cSS) et les autotransporteurs inversés (T5eSS). T6SS : Système de sécrétion protéique de Type VI. T7SS : Système de sécrétion protéique de Type VII ou voie chaperon-escorteur (CU). T8SS : système de sécrétion protéique de Type VIII ou voie de nucléation-précipitation extracellulaire (ENP). Cyt : cytoplasme. IM : membrane interne. Peri : périplasme. OM : membrane externe. EM : milieu extracellulaire.

Les *Escherichia coli* pathogènes intestinaux : Perspectives pour le développement de vaccins (Intestinal pathogenic *Escherichia coli*: insights for vaccines development)
Article de revue (soumis à Frontiers in Cellular and Infection Microbiology)

Maricarmen ROJAS-LOPEZ [1,2], Ricardo MONTERIO[2], Mariagrazia PIZZA [1], Mickaël DESVAUX [2,*] and Roberto ROSINI [1,*]
[1] GSK, Via Fiorentina 1, 53100 Siena, Italy, [2] Université Clermont Auvergne, INRA, UMR454 MEDiS, 63000 Clermont-Ferrand, France, * Auteurs correspondants.

Les maladies diarrhéiques sont l'une des principales causes de mortalité chez les enfants de moins de cinq ans et les *Escherichia coli* pathogènes intestinaux (InPEC) sont l'un des grands groupes responsables de ces infections dans le monde entier. À cet égard, les InPEC contribuent de manière significative aux maladies intestinales qui constituent un problème majeur dans les pays en voie de développement. Les pathotypes intestinaux tels que les *E. coli* entéropathogènes (EPEC) et les *E. coli* entérotoxinogènes (ETEC) sont principalement endémiques dans les pays en voie de développement, alors que les souches ETEC sont la principale cause de diarrhée (tourista) chez les voyageurs en visite dans ces pays. D'autre part, les *E. coli* entérohémorragiques (EHEC) sont à l'origine de grandes épidémies dans le monde, affectant principalement les pays développés et responsables non seulement de maladies diarrhéiques pouvant évoluer en colite hémorragique, mais aussi de complications cliniques sévères comme le syndrome hémolytique et urémique (SHU). Si on considère l'émergence de souches résistantes aux antibiotiques, l'augmentation annuelle des coûts du système de santé, l'incidence

élevée de la diarrhée des voyageurs et l'augmentation du nombre d'épisodes de SHU, la nécessité de développer des traitements préventifs efficaces s'avère cruciale.

Bien que l'utilisation d'antibiotiques soit toujours un facteur clé pour le traitement de ces infections, les stratégies non antibiotiques sont déterminantes pour limiter l'augmentation des résistances aux antibiotiques mais aussi absolument nécessaires dans le cas des infections à EHEC, pour lesquelles aucun traitement prophylactique n'est disponible et où l'utilisation d'antibiotiques n'est absolument pas recommandée. Dans ce contexte, le développement de vaccins se focalisant sur plusieurs déterminants moléculaires antigéniques extracellulaires et/ou de surface, dont des facteurs de virulence, apparaît comme une stratégie de choix (Figure 3). Cependant, à ce jour, aucun vaccin homologué n'est efficace contre les infections à InPEC. Ces dernières années, un effort important a été fourni pour développer des vaccins efficaces capables de prévenir les maladies diarrhéiques dues aux InPEC. Un certain nombre de jalons a déjà été posé qui offrent des perspectives prometteuses pour le développement de vaccins contre les InPEC.

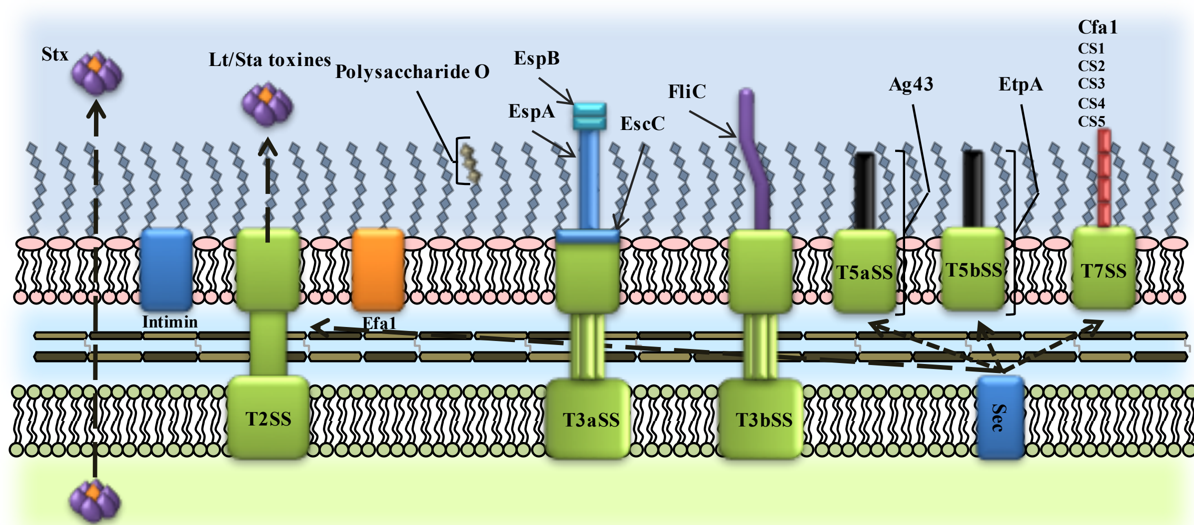


Figure 3 : **Représentation schématique des facteurs de virulence utilisés comme antigènes candidats pour le développement de vaccins chez les InPEC**. Différents facteurs de virulence ont été identifiés chez les InPECs comme des antigènes potentiels à des fins vaccinales, notamment des protéines de la membrane externe, des toxines, des polysaccharides O, des protéines exportés et sécrétées. T2SS : Système de sécrétion de Type II, T3aSS : Système de sécrétion de Type III sous-type a, T3bSS : Système de sécrétion de Type III sous-type b, T5aSS : Système de sécrétion de Type V sous-type a, T5bSS : Système de sécrétion de Type V sous-type b, T7SS : Système de sécrétion de Type VII.

RESULTATS DE LA THESE

L'augmentation de l'incidence des infections à EHEC ainsi que d'autres maladies diarrhéiques causées par différents agents étiologiques, est l'une des raisons pour poursuivre le développement et la découverte de nouveaux traitements contre ces pathogènes. De plus, l'utilisation d'antibiotiques n'est pas recommandée dans le cas d'infections à EHEC puisqu'elle peut aggraver la pathologie qui peut notamment évoluer vers des complications sévères, telle que le syndrome hémolytique et urémique (SHU). L'augmentation du nombre de SHU est devenue un problème majeur de santé publique, non seulement dans les pays occidentaux mais aussi en Amérique latine, comme en Argentine et au Brésil. À ce jour, il n'existe aucun vaccin homologué capable de protéger contre les *E. coli* pathogènes, que ce soit les ExPEC ou les InPEC, dont les EHEC. Les approches traditionnelles pour le développement de vaccins, en particulier l'atténuation de souches bactériennes pathogènes, nécessitent des années de travail qui n'aboutissent pas toujours à un vaccin efficace. Afin de tirer avantage des données « omiques », en particulier génomiques, transcriptomiques, ou encore protéomiques, de nouveaux concepts pour un développement plus rationnel et plus rapide de vaccins ont émergés. Ainsi, le concept de vaccinologie inverse a été formulé comme l'une des voies les plus prometteuses pour le développement des futurs vaccins. Cette approche s'appuie sur l'examen minutieux de toutes les CDS d'un génome, de prédire leur fonction et leur localisation subcellulaire. Selon les stratégies mises en œuvres, des données pan-génomiques, structurales pour la prédiction d'épitopes, transcriptomiques et/ou protéomiques peuvent être combinées selon différents pipelines d'analyse pour au final identifier des antigènes potentiels qui sont ensuite testés comme candidats au développement de nouveaux vaccins.

La sécrétion de facteurs de virulence spécifiques est l'une des principales différences entre les bactéries commensales et pathogènes. De tels facteurs de virulence sécrétés peuvent être libérés dans le milieu extracellulaire ou exposés à la surface cellulaire et être impliqués dans l'adhésion, la formation de biofilm et/ou la persistance des bactéries dans l'hôte. Le locus LEE code pour des déterminants impliqués dans la pathogénicité qui sont parmi les mieux caractérisés chez les EPEC et EHEC. Néanmoins, d'autres déterminants moléculaires peuvent participer à la pathogenèse et/ou au processus de colonisation depuis le portage animal, l'aliment comme vecteur de contamination jusqu'à l'infection chez les humains. Les autotransporteurs étant des

protéines pouvant se retrouver exposées à la surface bactérienne, ils peuvent être impliqués dans la reconnaissance immunologique et donc être de potentiels antigènes candidats pour le développement de vaccins. Au cours d'analyses immunoprotéomiques et dans une stratégie d'identification d'antigènes capables de générer une réponse immunitaire chez l'hôte au cours du processus d'infection, des autotransporteurs tels que EtpA, Ag43 et TibA ont pu être détectés. Ces autotransporteurs sont reconnus par les sérums de patients atteints de diarrhée à ETEC, confirmant leur expression au cours de l'infection. Ag43 et pAT sont capables d'augmenter le niveau IgA fécal et de fournir une protection relative contre la colonisation intestinale par les ETEC chez des souris immunisées. Par ailleurs, ils peuvent être impliqués dans différents processus de colonisation, tels que l'adhésion et la formation de biofilm.

Identification de nouveaux antigènes candidats chez *E. coli* O157:H7 par une stratégie de vaccinologie inverse

L'objectif de la première partie de cette thèse consiste à identifier de nouveaux antigènes candidats pour le développement de vaccins contre les EHEC O157:H7 par une approche de vaccinologie inverse. Au cours de ce travail, une approche bioinformatique d'analyse du génome de la souche EHEC O157: H7 EDL933 a été effectué afin d'identifier de nouveaux antigènes candidats en considérant la prédiction de leur fonction et localisation subcellulaire. Parallèlement à l'analyse de distribution de ces gènes, la prévalence et la variabilité des séquences des antigènes ont été analysées par rapport à d'autres souches et sérotypes d'EHEC, mais également d'autres *E. coli* pathogènes. Les antigènes sélectionnés ont ensuite été exprimés et délivrés selon la technologie GMMA afin d'évaluer leur capacité à conférer une immunoprotection et une réduction de la colonisation intestinale dans un modèle murin.

Identification d'un nouvel antigène candidat pour le développement de vaccins contre les *Escherichia coli* entérohémorragiques (EHEC) O157:H7 par une approche de vaccinologie inverse (Identification of a novel vaccine candidate against enterohemorrhagic *E. coli* (EHEC) O157:H7 by reverse vaccinology approach).
Article de recherche (prochainement soumis)

Maricarmen Rojas-Lopez[1,2], Manuele Martinelli[1], Valentina Brandi[3], Gregory Jubelin[2], Fabio Polticelli[3,4], Marco Soriani[1], Mariagrazia Pizza[1], Mickaël Desvaux[2], Roberto Rosini[1]*.
[1] GlaxoSmithKline Vaccines, Via Fiorentina 1, 53100, Siena, Italy, [2] Université Clermont

Auvergne, INRA, UMR454 MEDiS, 63000 Clermont-Ferrand, France, [3] UniROMA Tre, Department of Sciences, Viale G. Marconi 446, 00146 Rome, Italy, [4] National Institute of Nuclear Physics, Roma Tre Section, Rome, Italy. * Auteur correspondant.

Les *E. coli* entérohémorragiques (EHEC) sont une cause majeure de grandes épidémies affectant principalement les pays développés. De 1982 à 2002, un total de 350 épidémies impliquant des *E. coli* O157 ont été signalées aux Etats-Unis d'Amérique. Les infections par les EHEC provoquent des maladies diarrhéiques comme la colite hémorragique pouvant évoluer vers des complications cliniques sévères, comme le syndrome hémolytique et urémique (SHU). Bien que des efforts axés sur l'hygiène aient été mis en œuvre dans les chaînes d'approvisionnement alimentaire afin de réduire le risque d'infection par les *E. coli* O157 d'origine alimentaire, les épidémies causées par ce pathogène sont encore courantes et ce sans compter les cas sporadiques qui sont encore plus nombreux. Concernant le traitement des infections à EHEC, il faut noter que les thérapies à base d'antibiotiques sont généralement proscrites en raison du risque accru de libération de Shiga toxines (Stx), l'un des principaux facteurs de virulence de ces bactéries pathogènes. Parmi les stratégies de prévention non antibiotiques, le développement d'un vaccin est une approche pertinente pour les populations à risque mais à ce jour aucun vaccin homologué et spécifique contre les EHEC n'est disponible chez l'Homme. L'application d'une approche de vaccinologie inverse à partir du génome de la souche d'EHEC O157:H7 EDL933 a permis au final de sélectionner 3 antigènes protéiques potentiels (Figure 4). Ces antigènes ont été exprimés et délivrés par la technologie GMMA (modules généralisés pour les antigènes membranaires). Ces vaccins candidats se sont avérés immunogènes, et une réponse d'anticorps spécifique a pu être démontrée pour deux d'entre eux. De plus, l'immunisation avec l'antigène MC001 a permis de réduire significativement la colonisation intestinale par EHEC O157:H7 comme démontré par numération bactérienne dans les contenus fécaux, caecaux et au niveau du côlon chez la souris. Cet homologue de l'enzyme lipidase A de *Salmonella* Typhimurium (LpxR) est pour la première fois décrit comme un antigène candidat pour le développement de vaccins. La distribution des gènes et l'analyse de la variabilité des séquences ont montré que MC001 était principalement présent et conservé chez les EHEC O157:H7 et certains *E. coli* entéropathogènes (EPEC). Compte tenu de la forte variabilité génétique entre et au sein de ces entéropathotypes d'*E. coli*, l'utilisation d'un antigène aussi conservé que MC001 permettrait une couverture vaccinale élargie aux souches d'InPEC majeures.

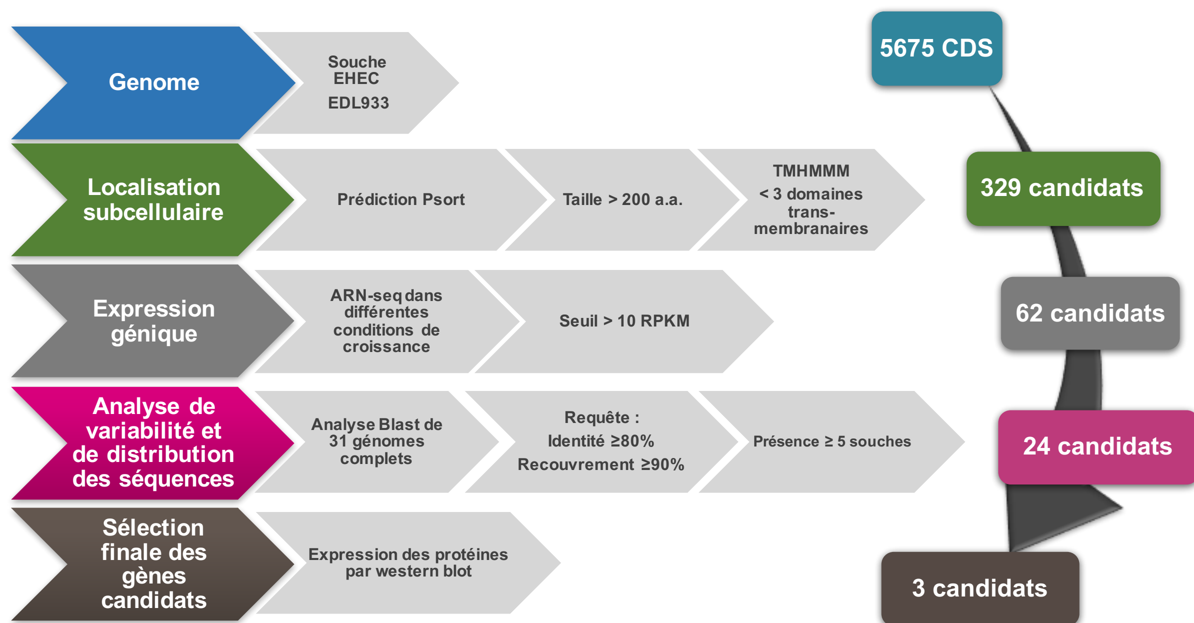


Figure 4 : **Description de la stratégie de vaccinologie inverse appliquée pour la sélection d'antigènes candidats à des fins vaccinales contre les EHEC O157:H7**. Les 5675 séquences d'ADN codant (CDS) encodées sur le chromosome de la souche *E. coli* O157:H7 EDL933 ont été analysés pour identifier des antigènes potentiels sur la base de leur localisation subcellulaire prédite par PSORT , ainsi que des critères de taille de la protéine et de nombre de domaines transmembranaires sur la base de prédiction par TMHMM. Des données sur l'expression génétique ont été intégrées au pipeline, en particulier issues d'analyses transcriptomiques par séquençage de l'ARN (ARN-seq) avec un seuil de 10 RPKM (lecture par kilobase par million de lectures cartographiés). La variabilité génétique et l'analyse de la distribution des antigènes potentiels a été effectué sur 31 génomes complets d'*E. coli* afin de sélectionner les antigènes candidats les plus conservés chez des InPEC. La sélection finale était également basée sur l'expression protéiques des antigènes candidats.

Identification du domaine autochaperon chez les protéines autotransporteurs du système de sécrétion de Type Va

L'approche de vaccinologie inverse précédemment appliquée a permis d'identifier MC021 en tant qu'antigène potentiel pour le développement de vaccins. Cet antigène appartient à la famille des autotransporteurs qui correspondent au système de sécrétion de type V sous-type a (T5aSS). Les autotransporetuṛs sont des protéines modulaires composées (i) d'un peptide signal N-terminal, qui est clivé après franchissement de la membrane interne, (ii) d'un domaine passager

central, souvent décrit avec un repliement en hélice β, (iii) d'une unité de translocation C-terminale, appelée translocateur, composée d'un domaine C-terminal en tonneau β, qui sert de pore pour la translocation au travers la membrane externe du domaine passager, et d'une région de liaison en hélice α qui relie le domaine passager au tonneau β. Le domaine passager, correspondant au domaine fonctionnel d'un autotransporteur, peut être exposé à la surface cellulaire ou libéré dans le milieu extracellulaire. Certains domaines passagers présentent également un domaine autochaperon impliqué dans la sécrétion du passager à travers la membrane externe et son repliement en surface. Cependant, la similitude et la distribution de ce domaine parmi les autotransporteurs restent controversées. En utilisant l'unique structure résolue d'un autochaperon chez IcsA, la prévalence de ce domaine a été explorée chez les autotransporteurs par modélisation structurale, analyses phylogénétiques et analyse des séquences protéiques. En clarifiant la prévalence de la région autochaperon au sein du T5aSS, il est apparu que ce domaine structural était uniquement associé aux domaines passagers avec une structure en hélice β soutenant l'idée que cette région pourrait être utilisée comme cible épitopique potentielle et/ou le développement de nouveaux traitements contre des autotransporteurs impliqués dans la virulence.

Identification du domaine autochaperon parmi les protéines du système de sécrétion de Type Va (T5aSS) : Caractéristique prédominante des autotransporteurs avec un domaine passager β-hélicoïdal (Identification of the autochaperone domain in the Type Va secretion system (T5aSS): A prevalent feature of autotransporters with a β–helical passenger).

Article de recherche (accepté dans Frontiers in Microbiology)

Maricarmen ROJAS-LOPEZ [1,2,#], Mohamed A. ZORGANI [1,#], Lawrence A. KELLEY [3], Xavier BAILLY [4], Andrey V. KAJAVA [5], Ian R. HENDERSON [6], Fabio POLTICELLI [7,8], Mariagrazia PIZZA [2], Roberto ROSINI [2] and Mickaël DESVAUX [1,*].

[1] Université Clermont Auvergne, INRA, UMR454 MEDiS, 63000 Clermont-Ferrand, [2] GlaxoSmithKline Vaccines, Via Fiorentina 1, 53100 Siena, Italy, [3] Imperial College London, Structural Bioinformatics Group, London, United Kingdom, [4] INRA, UR346 Epidémiologie animale, Saint Genès Champanelle, France, [5] Centre de Recherches en Biologie cellulaire de Montpellier, UMR5237 CNRS, Institut de Biologie Computationnelle, Université Montpellier, 34293 Montpellier, France, [6] University of Birmingham, Institute of Microbiology and Infection, Edgbaston, Birmingam, United Kingdom, [7] Department of Sciences, Roma Tre

University, Rome, Italy, [8]National Institute of Nuclear Physics, Roma Tre Section, Rome, Italy.
[#] Co-premier auteurs, * Auteur correspondant.

Les autotransporteurs (AT) appartiennent à une famille de protéines modulaires sécrétées par le système de sécrétion de Type V sous-type a (T5aSS) et considérées comme une source importante de facteurs de virulence chez les bactéries lipopolysaccharidiques didermes (bactéries à Gram négatif archétypales). Alors qu'ils sont exportés par la voie Sec, les AT sont sécrétés au travers la membrane externe via leur propre translocateur C-terminal formant un tonneau β, à travers lequel le reste de la protéine, à savoir le domaine passager, peut passer. Dans plusieurs AT, un domaine autochaperon (AC) présent dans la région C-terminale du domaine passager et en amont du translocateur a été démontré comme strictement nécessaire pour une sécrétion et un repliement corrects de ces protéines. Cependant, considérant qu'il était fonctionnellement caractérisé et identifié seulement dans une poignée d'ATs, une suspicion s'est récemment installée concernant la similitude et la conservation de cet élément structurel dans le T5aSS. Pour contourner le problème de la divergence des séquences primaires et tirer profit de la structure tridimensionnelle résolue de l'AC de l'AT IcsA, l'identification de ce domaine a été effectuée après alignement structural entre tous les domaines passagers d'AT résolus expérimentalement par cristallographie avant de rechercher ce domaine structural dans une banque de données de référence contenant 1523 ATs (Figure 5). Tout en démontrant que l'AC est en effet une structure conservée dans de nombreux ATs, l'analyse phylogénétique a révélé une distribution dans des branches profondément enracinées, d'où émergent 7 groupes principaux. L'analyse des séquences a révélé qu'un AC pouvait être identifié dans la grande majorité des ATs auto-associatifs mais pas dans les ATs lipases/estérases. L'analyse structurale a indiqué qu'un AC était présent chez les passagers présentant une hélice β, quel que soit le type d'enroulement β-solénoïde, mais pas avec les domaines passagers présentant un repliement globulaire α-hélicoïdal. Cette étude apporte un nouvel éclairage sur la biologie moléculaire de la sécrétion et du repliement des protéines via le T5aSS, où l'AC apparaît comme un élément structural prédominant et conservé exclusivement associé aux ATs présentant un domaine passager β-hélicoïdal.

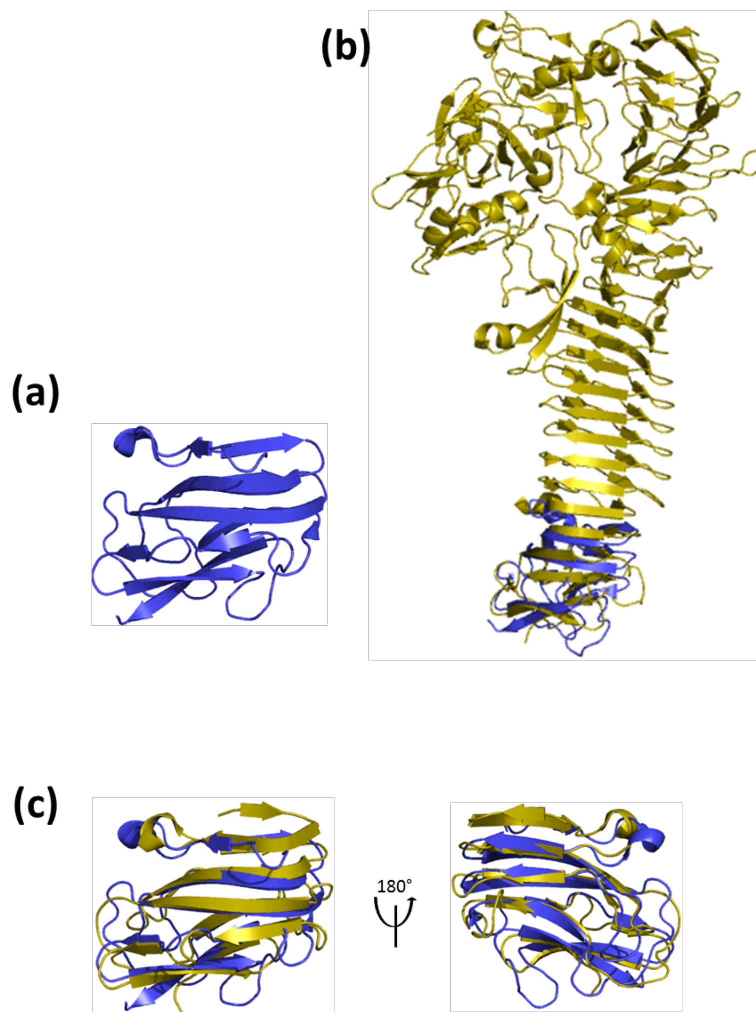


Figure 5 : **Identification structurale du domaine autochaperon (AC) dans des structures résolues de domaines passagers d'autotransporteurs**. (a) Domaine AC de IcsA raffiné (PDB: 3ML3; L606-D720). (b) Identification structurelle du domaine AC (bleu) dans la région C-terminale du passager de Esp (PDB: 3SZE, jaune). (c) Domaines AC de IcsA (D606-L720) et EspP (D869-A979) superposés.

CONCLUSION ET PERSPECTIVES

Au cours de ce travail, la mise en œuvre d'une stratégie de vaccinologie inverse combinée à la technologie GMMA nous a conduit à identifier et tester de nouveaux antigènes candidats. Cette approche nous a permis d'aller au-delà des antigènes typiques jusqu'à présent décrits à des fins vaccinales contre les EHEC, comme ceux codés par le LEE. Suite à ces travaux, la déacylase LpxR a pour la première fois été décrite comme un antigène candidat pour le développement de vaccins.

Tandis que notre approche de vaccinologie inverse tenait compte de prédictions liées aux fonctions et la localisation subcellulaire des CDS, des pipelines d'analyse plus poussés pourraient être développés. Par rapport à la localisation subcellulaire des protéines, l'application du concept de sécrétome permettrait d'améliorer ce type de prédiction en prenant en compte la biologie de la sécrétion. En effet, des outils de prédictions de la localisation finale des protéines comme PSORT ne considèrent pas les voies de sécrétion et de maturation des protéines. La combinaison de plusieurs outils de prédiction décomposant la sécrétion de protéines depuis l'exportation dans le périplasme à la sécrétion au travers de la membrane externe, ainsi que les voies de maturation post-translocationnelle où les domaines d'ancrage ou de rétention aux membranes sont considérées comme une approche plus pertinente et puissante pour ce type de prédiction. Malheureusement, ce type de pipeline d'analyse n'a pas encore été développé pour des bactéries lipopolysaccharidiques didermes. Néanmoins, cette stratégie pourrait s'avérer limitée par certains mécanismes moléculaires impliqués dans l'adressage et la sécrétion qui restent encore mal connus ; par exemple, les mécanismes d'adressage des protéines exportées dans le périplasme vers le T2SS ou des protéines vers le T3SS, qui par conséquent ne peuvent pas être prédit sans a priori. De plus, l'existence d'erreurs dans les annotations de génomes peut conduire à l'exclusion d'antigènes potentiels, ce qui peut nécessiter de revoir l'identification de CDS par des approches bioinformatiques plus récentes. L'inclusion d'autres données « omiques », comme des données transcriptomiques (ARN-seq), protéomiques (sous-protéomes) ou encore métabolomiques, fournit des informations supplémentaires sur l'expression des gènes, qui est un point crucial pour l'efficacité d'un vaccin. En effet, l'étude de l'expression *in fine* protéique dans différentes conditions environnementales, en particulier par rapport à l'hôte, aide à cibler les antigènes les plus pertinents. Plus récemment, le concept de vaccinologie inverse 2.0 vise à affiner la découverte de nouveaux vaccins en intégrant la caractérisation structurale et épitopique des antigènes combinés aux nouvelles découvertes sur les anticorps humains protecteurs (Rappuoli et al., 2016).

La vaccinologie inverse cible essentiellement des antigènes exposés en surface qui sont ainsi soit localisés à la membrane externe, sécrétés dans le milieu extracellulaire ou présents dans des organelles de surface. De ce point de vue, les autotransporteurs apparaissent comme une cible de choix et ont souvent été utilisés dans le développement de vaccins contre d'autres InPEC, notamment les ETEC. Afin d'identifier des épitopes potentiels sur ces protéines, l'analyse des structures tridimensionnelles est une approche pertinente pour étudier les régions exposées et leur conformation. Les domaines autochaperons, que nous avons démontrés comme

systématiquement associés à des domaines passagers avec un repliement en hélice β, sont importants pour le repliement et la sécrétion des autotransporteurs. D'un point de vue biotechnologique, ces domaines sont donc importants pour améliorer l'expression de protéines, notamment hétérologues, via le T5aSS, comme par exemple pour l'exposition de protéines de surface sous une conformation active et fonctionnelle, ou d'un point de vue biomédical, pour bloquer la sécrétion et/ou le repliement correct d'autotransporteurs impliqués dans la pathogénicité et ainsi atténuer les niveaux de virulence de bactéries pathogènes ou pour l'exposition d'épitopes cibles dans une stratégie vaccinale. Dans ce contexte, l'utilisation combinée du système GMMA et de domaines autochaperons qui peuvent agir en *trans* est une direction de recherche qui mériterait d'être explorée et investiguée de manière approfondie.

Par rapport aux 24 gènes candidats identifiés suite à notre analyse bioinformatique certains pourraient être caractérisés afin d'élucider leurs propriétés et fonctions, notamment leur rôle dans la pathogenèse. De même, l'étude des propriétés immunitaires *in vitro* et *in vivo* serait nécessaire afin de caractériser les potentialités de ces candidats comme antigènes pour le développement du vaccin. Certains antigènes peuvent se trouver exprimés et purifiés sous une forme insoluble ; afin, d'améliorer la présentation des épitopes, l'approche classique consiste à tester d'autres conditions d'expression et de purification tandis que les dernières avancées dans le domaine tendent à privilégier l'approche GMMA. Outre le système GMMA qui est une application technologique des vésicules membranaires bactériennes pour le développement de vaccins, l'implication des OMVs (Outer Membrane Vesicles) dans la physiopathologie des EHEC et des InPEC en générale est une direction de recherche qui n'a encore été que peu explorée. Sans doute l'une des premières directions de recherche consisterait en l'analyse protéomique de ces OMVs chez les différents pathotypes d'InPEC, à partir des souches prototypiques, afin d'établir leurs répertoires protéiques sous différentes conditions environnementales. De plus, l'implication de ces OMVs dans la physiopathologie des InPEC et des EHEC en particulier est une question pertinente, par rapport à leur rôle dans la fixation ou l'évasion de la réponse immunitaire ou encore dans le processus de colonisation intestinale ainsi que l'adhésion et la formation de biofilm (Kulp et al., 2010, Schwechheimer et al., 2015).

Outre MC007, il serait nécessaire de caractériser MC020 plus avant. L'autotransporteur MC020 semble avoir déclenché une réponse immunitaire avec la production d'anticorps IgG spécifiques. Concernant MC001, des questions spécifiques restent à élucider, comme savoir si l'effet de GMMA-MC001 est directement attribué à LpxR ou s'il résulte de l'activité déacylase, comme

par exemple des modifications du LPS. De plus, des études supplémentaires seront nécessaires pour caractériser l'implication potentielle de LpxR dans la virulence et la pathogénicité des EHEC. Certaines de ces analyses peuvent inclure des études de déacylation du LPS, des tests pour définir son rôle dans la physiopathologie, sa capacité à réguler la réponse immunitaire, ou encore les interactions avec l'hôte et le microbiote. Par ailleurs et afin d'améliorer la réponse immunitaire et la protection vaccinale, d'autres voies d'immunisation pourraient être testées. La préparation du vaccin nécessitera également des études sur la formulation avec différents adjuvants afin de déterminer le plus adéquat ainsi que l'optimisation d'une production à grande échelle.

L'espèce *E. coli* est bien connue pour sa très haute plasticité génétique et génomique comme l'indique notamment la présence de souches commensales et d'une grande diversité de souches pathogènes. Cette plasticité pose également problème pour le développement de traitements efficaces face à la diversité des maladies causées par cette bactérie. Néanmoins, les souches d'*E. coli* partagent un même noyau génomique, c'est-à-dire un patrimoine génétique commun et très conservé. Les gènes codant pour des facteurs de virulence, c'est-à-dire impliqués dans les fonctions pathogènes responsables de l'infection, le développement de la maladie et l'interaction hôte-pathogène, varient selon les pathotypes et peuvent être localisés au sein de loci, appelés îlots de pathogénicité, comme le LEE. Le séquençage de génomes entiers permet de décrypter les variations clonales de populations bactériennes de manière beaucoup plus détaillée que les approches de typage moléculaire multi-locus (MLST) et ouvre la voie à de nouvelles perspectives sur l'évolution des souches pathogènes, qui devrait être à l'avenir intégré de manière plus systématique à des approches de vaccinologie inverse.

Les efforts pour prévenir les maladies liées à des infections par les *E. coli* pathogènes sont essentiellement focalisés sur les UPEC chez les ExpEC et sur les ETEC chez les InPEC. Tandis que ce travail s'est focalisé sur les EHEC O157, la prévention de maladies diarrhégéniques nécessiterait d'étendre ce type d'étude aux autres entéropathotypes d'*E. coli* qui posent eux aussi des problèmes de santé publique et qui peuvent se retrouver dans des cas de co-infections, encore trop peu considérés et envisagés dans les cas d'infections. Ces autres InPEC comprennent d'une part les sérogroupes d'EHEC non-O157 et d'autre part les EPEC et EAEC mais aussi les DAEC ou encore EIEC, dont les *Shigella*, pour lesquels les recherches menées pour le développement de traitements curatifs ou préventifs, tels que le développement de vaccins, sont rares ou inexistantes (Bouzari et al., 2010). Outre le sérotype O157: H7, les

infections à EHEC peuvent être associées à 6 autres sérotypes, nommément O26:H11, O45:H2, O103:H11, O111:H^-, O121:H19, O145:H^-.

Outre le développement de nouvelles approches de vaccinologie inverse pour la découverte de nouveaux antigènes (García-Angulo et al., 2014), cette démarche doit de nos jours s'insérer au concept One Health pour une recherche multidisciplinaire et intégrée à la chaîne alimentaire (Torres, 2017). En effet, ce concept stipule et rappelle que la santé humaine est directement liée aux relations entre l'environnement, les animaux et les êtres humains. Pour cette raison, les recherches ne doivent pas rester focalisées sur la seule problématique de l'infection de ces bactéries pathogènes chez l'hôte, en l'occurrence l'Homme, mais incorporer la dimension écophysiologique. Ceci nécessite d'étudier ces bactéries pathogènes aussi chez l'animal et dans d'autres environnements, qu'ils soient naturels, alimentaires ou industriels, afin de pleinement intégrer la notion de réservoir animal, de passages de pathogènes comme les InpEC entre le tractus digestif et ces environnements, ce qui est d'autant plus pertinent dans le cas de pathogènes alimentaires zoonotiques (Etcheverría et al., 2016). La lutte contre les infections à InPEC est donc toujours en cours et les perspectives de développement de nouvelles mesures de prévention ou de traitements curatifs ou préventifs nécessitent une recherche multidisciplinaire pour obtenir des avancées significatives. Des efforts supplémentaires, non seulement au niveau de la recherche biomédicale ou fondamentale mais aussi dans les domaines clinique, vétérinaire et de la santé publique, sont nécessaires pour mieux comprendre la prolifération et l'émergence de nouvelles maladies liées aux *E. coli*, comme les souches hybrides EHEC/EAEC lors des dernières épidémies en Europe ou encore les *E. coli* adhérents invasifs (AIEC) associés à la maladie de Crohn.

Suplementary Material

Second part: Results

Chapter 1
Identification of a novel vaccine candidate against Enterohemorrhagic *E. coli* (EHEC) O157:H7 by Reverse Vaccinology approach

&

Chapter 2
Identification of the autochaperone domain in the Type Va secretion system (T5aSS): A prevalent feature of autotransporters with a β–helical passenger

1.- Table 1S: Identification of AC domains in the well-defined dataset of 1523 autotransporters (Celik *et al.*, 2012).

2.- PDB files restricted to the AC domains identified in the resolved structure of the passenger of AC^{IcsA} (PDB: 3ML3; D_{606}-L_{720}), AC^{EspP} (PDB: 3SZE; D_{869}–I_{979}), AC^{Hbp} (PDB: 1WXR; N_{948}-L_{1056}), AC^{Pet} (PDB: 4OM9; N_{865}-I_{974}), AC^{P69} (PDB: 1DAB; D_{444}-L_{556}), AC^{Hap} (PDB: 3SYJ; D_{830}-L_{964}) and AC^{IgA1} (PDB: 3H09; D_{865}-L_{977}).

Chapter 1
Identification of a novel vaccine candidate against Enterohemorrhagic *E. coli* (EHEC) O157:H7 by Reverse Vaccinology approach

Table S1. Vaccine candidates selected by Psort analysis

Protein ID	Functional annotation	Length (bp)	PSORT Localization	PSORT score	ModHMM
AIG68144.1	Ferric siderophore transport system, periplasmicbinding protein TonB CDS	720	Unknown	2	1 internal helix
AIG70214.1	Type III secretion outermembrane pore forming protein (YscC,MxiD,HrcC, InvG) CDS	1704	OuterMembrane	10	1 internal helix
AIG71811.1	adherence and invasion outermembrane protein (Inv,enhances Peyer's patches colonization) CDS	5037	OuterMembrane	9.95	No internal helices
AIG66227.1	Putative outer membrane protein CDS	2451	OuterMembrane	10	No internal helices
AIG66265.1	Outer membrane protein Imp, required for envelope biogenesis CDS	2322	OuterMembrane	10	No internal helices
AIG66267.1	hypothetical protein CDS	753	Unknown	2	No internal helices
AIG66308.1	Secretion monitor precursor CDS	588	Unknown	6.49	1 internal helix
AIG66347.1	Fimbrial protein Yad like protein CDS	1110	Unknown	2	1 internal helix
AIG66348.1	Fimbrial protein YadK CDS	591	Unknown	2	No internal helices
AIG66349.1	Fimbrial protein YadL CDS	606	Unknown	2.5	1 internal helix
AIG66351.1	Outer membrane usher protein HtrE CDS	2601	OuterMembrane	10	No internal helices
AIG66353.1	Fimbrial protein YadN CDS	597	Extracellular	9.65	1 internal helix
AIG66363.1	Ferric hydroxamate outer membrane receptor FhuA CDS	2244	OuterMembrane	10	No internal helices
AIG66390.1	Outer membrane protein assembly factor YaeT precursor CDS	2433	OuterMembrane	10	1 internal helix
AIG66406.1	Copper homeostasis protein CutF precursor CDS	711	OuterMembrane	9.92	No internal helices
AIG66407.1	putative lipoprotein yaeF precursor CDS	825	Unknown	2.5	No internal helices
AIG66409.1	hypothetical protein CDS	708	Unknown	2	No internal helices
AIG66417.1	hypothetical protein CDS	801	Unknown	2	No internal helices
AIG66424.1	putative aminopeptidase CDS	801	Unknown	2.5	1 internal helix
AIG66428.1	hypothetical protein CDS	1410	Unknown	2	1 internal helix
AIG66446.1	core protein CDS	4215	Unknown	2	No internal helices
AIG66447.1	hypothetical protein CDS	612	Unknown	2	1 internal helix
AIG66450.1	core protein CDS	1761	Extracellular	9.52	No internal helices
AIG66460.1	putative exported protein CDS	741	Unknown	2.5	1 internal helix
AIG66464.1	putative lipoprotein yafL precursor CDS	774	Unknown	4.9	1 internal helix
AIG66480.1	Outer membrane pore protein E precursor CDS	1056	OuterMembrane	10	1 internal helix
AIG66490.1	putative tail fiber protein CDS	795	Unknown	2	No internal helices
AIG66511.1	Zinc binding domain protein CDS	2334	Unknown	2	No internal helices
AIG66518.1	hypothetical protein CDS	1137	Unknown	2.5	1 internal helix
AIG66526.1	CFA/I fimbrial chaperone CDS	711	Unknown	2.5	1 internal helix
AIG66527.1	CFA/I fimbrial minor adhesin CDS	1644	Extracellular	9.64	1 internal helix
AIG66528.1	CFA/I fimbrial subunit C usher protein CDS	2526	Unknown	4.69	2 internal helices
AIG66528.1	Putative adhesin CDS	4254	OuterMembrane	10	No internal helices
AIG66545.1	AidA-I adhesin-like protein CDS	4050	Unknown	5.87	No internal helices
AIG66568.1	hypothetical protein CDS	864	Unknown	2	No internal helices
AIG66574.1	Putative flagellin structural protein CDS	2943	OuterMembrane	10	No internal helices
AIG66636.1	hypothetical protein CDS	624	Unknown	2	No internal helices
AIG66640.1	putative lipoprotein CDS	1095	Unknown	2.5	No internal helices

AIG66656.1	Protein YkiA CDS	2166	Unknown	2	No internal helices
AIG66683.1	Nucleoside-specific channel-forming protein Tsx precursor CDS	885	OuterMembrane	10	No internal helices
AIG66684.1	putative lipoprotein yajI CDS	540	Unknown	2	No internal helices
AIG66769.1	Large repetitive protein CDS	4386	Unknown	6.04	1 internal helix
AIG66781.1	Putative stomatin/prohibitin-family membrane protease subunit YbbK	918	Unknown	2	Unknown
AIG66819.1	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA CDS	1701	Unknown	2	No internal helices
AIG66819.1	Outer membrane usher protein SfmD CDS	2610	OuterMembrane	10	2 internal helices
AIG66854.1	Agglutination protein CDS	1356	OuterMembrane	10	1 internal helix
AIG66854.1	TonB-dependent receptor CDS	2241	OuterMembrane	10	No internal helices
AIG66909.1	Rare lipoprotein A precursor CDS	1089	Extracellular	9.65	No internal helices
AIG66920.1	hypothetical protein CDS	588	Extracellular	9.71	No internal helices
AIG66922.1	hypothetical protein CDS	708	Unknown	2	No internal helices
AIG66972.1	core protein CDS	4200	Unknown	2	No internal helices
AIG66973.1	orf, hypothetical protein CDS	543	Unknown	2	No internal helices
AIG66984.1	hypothetical protein CDS	1062	Unknown	2.5	No internal helices
AIG66987.1	putative fimbrial-like protein ygiL precursor CDS	567	Extracellular	9.72	No internal helices
AIG67020.1	TolA protein CDS	1185	Unknown	2	2 internal helices
AIG67060.1	Putative outer membrane protein CDS	960	Unknown	2.5	1 internal helix
AIG67099.1	Biotin synthesis protein BioC CDS	756	Unknown	2	No internal helices
AIG67129.1	Ferrichrome-iron receptor CDS	2283	OuterMembrane	10	No internal helices
AIG67212.1	Virulence factor VirK CDS	993	Unknown	2	No internal helices
AIG67245.1	Urease accessory protein UreF CDS	675	Unknown	2	No internal helices
AIG67247.1	Per-activated serine protease autotransporter enterotoxin EspC	126	Extracellular	9.71	Unknown
AIG67280.1	Colicin I receptor precursor CDS	2010	OuterMembrane	10	No internal helices
AIG67308.1	putative hemolysin activator protein CDS	1065	OuterMembrane	8.86	1 internal helix
AIG67367.1	Putative metalloprotease yggG	789	Unknown	2	Unknown
AIG67384.1	exported protein CDS	549	Unknown	2.5	1 internal helix
AIG67387.1	Outer membrane protein F precursor CDS	1089	OuterMembrane	10	1 internal helix
AIG67397.1	type 1 fimbriae major subunit FimA CDS	540	Extracellular	10	No internal helices
AIG67400.1	type 1 fimbriae anchoring protein FimD CDS	1893	OuterMembrane	10	No internal helices
AIG67407.1	hypothetical protein CDS	543	Unknown	2	No internal helices
AIG67412.1	Paraquat-inducible protein B CDS	1641	Unknown	2	1 internal helix
AIG67413.1	Paraquat-inducible protein B CDS	564	Unknown	2	No internal helices
AIG67419.1	Outer membrane protein A precursor CDS	1065	OuterMembrane	10	No internal helices
AIG67426.1	UPF0319 protein YccT precursor CDS	663	Unknown	2.5	1 internal helix
AIG67448.1	hypothetical protein CDS	558	Unknown	2	No internal helices
AIG67453.1	hypothetical protein CDS	774	Unknown	2	No internal helices
AIG67464.1	Hypothetical protein CDS	1851	Unknown	2	No internal helices
AIG67503.1	hypothetical protein CDS	753	Extracellular	9.71	No internal helices
AIG67516.1	Putative polysaccharide export protein YccZ precursor CDS	1140	OuterMembrane	9.92	1 internal helix
AIG67517.1	Putative outer membrane lipoprotein YmcA CDS	2097	OuterMembrane	9.52	No internal helices

AIG67577.1	Shiga-like toxin II subunit A precursor CDS	960	Unknown	2.5	1 internal helix
AIG67607.1	hypothetical protein CDS	618	Unknown	2	No internal helices
AIG67611.1	hypothetical protein CDS	657	Unknown	2	No internal helices
AIG67633.1	hypothetical protein CDS	1344	Extracellular	9.64	No internal helices
AIG67644.1	Biofilm PGA synthesis deacetylase PgaB CDS	2019	Unknown	2	1 internal helix
AIG67645.1	Biofilm PGA outer membrane secretin PgaA CDS	2424	Unknown	2.5	No internal helices
AIG67652.1	Putative exported protein precursor CDS	1086	Unknown	2	No internal helices
AIG67653.1	outer membrane fimbrial usher protein CDS	2523	OuterMembrane	10	No internal helices
AIG67658.1	Hemolysin CDS	3813	OuterMembrane	9.95	No internal helices
AIG67671.1	hypothetical protein CDS	789	Unknown	2	No internal helices
AIG67672.1	hypothetical protein CDS	1389	Unknown	2	No internal helices
AIG67699.1	Per-activated serine protease autotransporter enterotoxin EspC	126	Extracellular	9.71	Unknown
AIG67732.1	TonB-dependent receptor CDS	2010	OuterMembrane	10	No internal helices
AIG67760.1	putative hemolysin activator protein CDS	1065	OuterMembrane	8.86	1 internal helix
AIG67820.1	Protein yceI precursor CDS	576	Unknown	2.5	1 internal helix
AIG67827.1	Putative lipoprotein yceB precursor CDS	561	Unknown	2	No internal helices
AIG67831.1	Protein of unknown function YceH CDS	648	Unknown	2	No internal helices
AIG67832.1	Virulence factor MviM CDS	924	Unknown	2	No internal helices
AIG67840.1	Flagellar basal-body rod modification protein FlgD CDS	696	Extracellular	10	No internal helices
AIG67841.1	Flagellar hook protein FlgE CDS	1206	Extracellular	10	No internal helices
AIG67843.1	Flagellar basal-body rod protein FlgG CDS	783	Extracellular	10	No internal helices
AIG67844.1	Flagellar L-ring protein FlgH CDS	699	OuterMembrane	10	1 internal helix
AIG67847.1	Flagellar hook-associated protein FlgK CDS	1644	Extracellular	9.96	No internal helices
AIG67848.1	Flagellar hook-associated protein FlgL CDS	954	Extracellular	10	No internal helices
AIG67862.1	YceG like protein CDS	1023	Unknown	6.49	1 internal helix
AIG67867.1	Putative OMR family iron-siderophore receptor precursor CDS	2190	OuterMembrane	10	No internal helices
AIG67871.1	Lipoprotein YcfM, part of a salvage pathway of unknown substrate CDS	642	Unknown	2	No internal helices
AIG67876.1	Putative exported protein CDS	540	Unknown	2.5	1 internal helix
AIG67905.1	hypothetical protein CDS	588	Unknown	2	No internal helices
AIG67914.1	Hypothetical protein CDS	1938	Unknown	2	No internal helices
AIG67929.1	hypothetical protein CDS	1938	Unknown	2	No internal helices
AIG67963.1	hypothetical protein CDS	744	Unknown	2	No internal helices
AIG68016.1	hypothetical protein CDS	1602	Unknown	2	No internal helices
AIG68033.1	Attachment invasion locus protein precursor CDS	600	OuterMembrane	10	1 internal helix
AIG68048.1	Protease VII (Omptin) precursor CDS	954	OuterMembrane	10	No internal helices
AIG68053.1	Pertactin precursor CDS	705	Unknown	7	No internal helices
AIG68064.1	Hemolysin E, chromosomal CDS	1056	Extracellular	10	No internal helices
AIG68081.1	Putative TonB dependent outer membrane receptor CDS	1971	OuterMembrane	10	No internal helices
AIG68093.1	Putative adhesion and penetration protein CDS	1644	Extracellular	9.65	No internal helices
AIG68114.1	Invasin CDS	1431	OuterMembrane	10	No internal helices
AIG68125.1	UPF0028 protein YchK CDS	945	Extracellular	8.89	No internal helices

AIG68148.1	Outer membrane protein W precursor CDS	639	OuterMembrane	10	No internal helices
AIG68163.1	hypothetical protein CDS	582	Unknown	2	No internal helices
AIG68165.1	Putative intestinal colonization factor encoded by prophage CP-933O CDS	747	Unknown	2	No internal helices
AIG68173.1	Hypothetical protein CDS	1854	Unknown	2	No internal helices
AIG68183.1	hypothetical protein CDS	1938	Unknown	2	No internal helices
AIG68190.1	putative major tail subunit CDS	786	Extracellular	9.64	No internal helices
AIG68238.1	Hypothetical protein CDS	1851	Unknown	2	No internal helices
AIG68271.1	Attachment invasion locus protein precursor CDS	600	OuterMembrane	10	1 internal helix
AIG68309.1	Autoinducer 2 (AI-2) aldolase LsrF CDS	876	Unknown	2	No internal helices
AIG68316.1	hypothetical protein CDS	3894	Unknown	5.87	No internal helices
AIG68317.1	hypothetical protein CDS	1401	OuterMembrane	9.83	No internal helices
AIG68320.1	type 1 fimbriae major subunit FimA CDS	564	Extracellular	10	1 internal helix
AIG68322.1	type 1 fimbriae anchoring protein FimD CDS	2223	OuterMembrane	10	No internal helices
AIG68325.1	mannose-specific adhesin FimH CDS	915	Unknown	2	1 internal helix
AIG68333.1	hypothetical protein CDS	2373	OuterMembrane	9.49	No internal helices
AIG68337.1	redicted glycoside hydrolase CDS	1320	Unknown	2	No internal helices
AIG68357.1	Outer membrane porin protein NmpC precursor CDS	1101	OuterMembrane	10	1 internal helix
AIG68359.1	hypothetical protein CDS	669	Unknown	2	No internal helices
AIG68376.1	internalin, putative CDS	1260	Unknown	2	No internal helices
AIG68383.1	hypothetical protein CDS	1062	Unknown	2.5	No internal helices
AIG68385.1	core protein CDS	4203	Unknown	5.48	No internal helices
AIG68385.1	putative tonB-dependent receptor yncD precursor CDS	2019	OuterMembrane	10	No internal helices
AIG68406.1	putative membrane lipoprotein clustered with tellurite resistance proteins TehA/TehB CDS	669	Unknown	2	No internal helices
AIG68433.1	putative BigA-like protein CDS	2460	Unknown	6.26	No internal helices
AIG68444.1	Outer membrane protein N precursor CDS	753	OuterMembrane	10	No internal helices
AIG68484.1	porin, autotransporter (AT) family CDS	3036	Extracellular	9.65	2 internal helices
AIG68484.1	Hypothetical protein CDS	1854	Unknown	2	No internal helices
AIG68547.1	Outer membrane protein G precursor CDS	906	OuterMembrane	10	No internal helices
AIG68579.1	RND efflux system, outer membrane lipoprotein CmeC CDS	1374	OuterMembrane	9.98	No internal helices
AIG68638.1	Attachment invasion locus protein precursor CDS	600	OuterMembrane	10	1 internal helix
AIG68679.1	Hypothetical protein CDS	1851	Unknown	2	No internal helices
AIG68688.1	hypothetical protein CDS	585	Unknown	2	No internal helices
AIG68723.1	Putative protease ydgD	822	Unknown	2.5	Unknown
AIG68730.1	Protein ydgH precursor CDS	945	Unknown	2.5	No internal helices
AIG68784.1	Putative lipoprotein CDS	816	Unknown	6.49	No internal helices
AIG68794.1	putative enzyme CDS	1257	Unknown	7	No internal helices
AIG68797.1	hypothetical protein CDS	813	Unknown	2	No internal helices
AIG68812.1	Iron-sulfur cluster assembly protein SufB CDS	1488	Unknown	2	No internal helices
AIG68838.1	hypothetical protein CDS	714	Unknown	2	No internal helices
AIG68851.1	hypothetical protein CDS	1899	Unknown	2	No internal helices
AIG68853.1	Putative outer membrane protein CDS	759	Unknown	2.5	No internal helices

AIG68861.1	hypothetical protein CDS	768	Unknown	2	No internal helices
AIG68883.1	Hypothetical protein YdjY CDS	678	Unknown	2.5	No internal helices
AIG68886.1	ABC transporter, periplasmic substrate-binding protein YnjB CDS	1167	Unknown	2.5	No internal helices
AIG68897.1	Protein ydjA CDS	552	Unknown	2	1 internal helix
AIG68899.1	putative lipoprotein CDS	711	Unknown	2	1 internal helix
AIG68918.1	MltA-interacting protein MipA CDS	747	OuterMembrane	10	No internal helices
AIG68945.1	Starvation lipoprotein Slp-like protein CDS	582	OuterMembrane	9.92	1 internal helix
AIG68946.1	Inactive metal-dependent proteases like protein,putative molecular chaperone	696	Unknown	2	Unknown
AIG68978.1	Paraquat-inducible protein B CDS	2640	Unknown	2	2 internal helices
AIG68989.1	Protease II	2061	Unknown	5.41	Unknown
AIG69001.1	Cell wall endopeptidase, family M23/M37 CDS	1323	Unknown	2	1 internal helix
AIG69007.1	hypothetical protein CDS	603	Unknown	2	No internal helices
AIG69057.1	putative membrane protein CDS	651	Unknown	2	1 internal helix
AIG69095.1	Flagellar biosynthesis protein FliZ CDS	552	Unknown	2	No internal helices
AIG69098.1	Flagellar biosynthesis protein FliC CDS	1758	Extracellular	9.96	No internal helices
AIG69099.1	Flagellar hook-associated protein FliD CDS	1398	Extracellular	10	No internal helices
AIG69108.1	invasion plasmid antigen CDS	1140	Extracellular	8.89	No internal helices
AIG69110.1	invasion plasmid antigen CDS	1140	Extracellular	8.89	No internal helices
AIG69118.1	Flagellar hook-length control protein FliK CDS	1128	Extracellular	10	No internal helices
AIG69140.1	Outer membrane protein N precursor CDS	645	OuterMembrane	10	1 internal helix
AIG69141.1	Outer membrane protein N precursor CDS	576	OuterMembrane	10	No internal helices
AIG69149.1	putative zinc-binding lipoprotein ZinT CDS	651	Unknown	2.5	No internal helices
AIG69153.1	hypothetical protein CDS	711	Unknown	2	No internal helices
AIG69158.1	Attachment invasion locus protein precursor CDS	600	OuterMembrane	10	1 internal helix
AIG69189.1	Hypothetical protein CDS	1851	Unknown	2	No internal helices
AIG69198.1	hypothetical protein CDS	564	Unknown	2	No internal helices
AIG69216.1	adherence and invasion outermembrane protein (Inv,enhances Peyer's patches colonization) CDS	7863	OuterMembrane	9.95	No internal helices
AIG69305.1	AsmA protein CDS	1854	Unknown	2.5	1 internal helix
AIG69312.1	Putative chaperonin CDS	1941	Unknown	2	2 internal helices
AIG69314.1	hypothetical protein CDS	660	Unknown	2	No internal helices
AIG69346.1	Uncharacterized protein YehA precursor CDS	1035	Unknown	2	No internal helices
AIG69347.1	Fimbriae usher protein StcC CDS	2481	OuterMembrane	10	No internal helices
AIG69349.1	Putative fimbrial-like protein CDS	543	Unknown	2.5	1 internal helix
AIG69361.1	hypothetical protein CDS	1530	Unknown	2	No internal helices
AIG69362.1	hypothetical protein CDS	732	Unknown	2	No internal helices
AIG69377.1	Attachment invasion locus protein precursor CDS	600	OuterMembrane	10	1 internal helix
AIG69390.1	Minor tail protein Z CDS	624	Unknown	2	No internal helices
AIG69393.1	Prophage Clp protease-like protein	1041	Unknown		Unknown
AIG69411.1	Shiga toxin A-chain precursor CDS	948	Unknown	2.5	1 internal helix
AIG69429.1	putative superinfection exclusion protein CDS	555	Unknown	2	2 internal helices
AIG69485.1	Colicin I receptor precursor CDS	1884	OuterMembrane	10	No internal helices

AIG69508.1	Lipoprotein spr precursor CDS	567	Unknown	2.5	No internal helices
AIG69510.1	ABC transporter, periplasmic substrate-binding protein CDS	1815	Unknown	6.58	No internal helices
AIG69523.1	Putative ATP-binding component of a transport system CDS	2592	Extracellular	9.46	No internal helices
AIG69547.1	Outer membrane protein C precursor CDS	1104	OuterMembrane	10	1 internal helix
AIG69552.1	hypothetical protein CDS	777	Unknown	2	No internal helices
AIG69553.1	Putative membrane protein CDS	4515	Unknown	4.69	No internal helices
AIG69558.1	Type V secretory pathway, adhesin AidA CDS	3705	OuterMembrane	9.83	No internal helices
AIG69581.1	Polymyxin resistance protein PmrG CDS	603	Unknown	2.5	1 internal helix
AIG69644.1	DedD protein CDS	663	Unknown	2	1 internal helix
AIG69664.1	Uncharacterized protein YadU in stf fimbrial cluster CDS	843	Unknown	2	No internal helices
AIG69669.1	Fimbriae usher protein StfC CDS	2583	OuterMembrane	10	No internal helices
AIG69676.1	Long-chain fatty acid transport protein CDS	1341	OuterMembrane	10	No internal helices
AIG69679.1	Lipoprotein CDS	756	OuterMembrane	9.92	2 internal helices
AIG69706.1	hypothetical protein CDS	636	Unknown	2.5	No internal helices
AIG69727.1	putative virulence protein CDS	1149	Unknown	2	No internal helices
AIG69790.1	YpfJ protein, zinc metalloprotease superfamily	864	Unknown	2	Unknown
AIG69792.1	Outer membrane protein NlpB, lipoprotein component of the protein assembly complex (forms a complex with YaeT, YfiO, and YfgL) CDS	1035	OuterMembrane	9.93	No internal helices
AIG69809.1	Exported zinc metalloprotease YfgC precursor	1464	Unknown	2.5	Unknown
AIG69829.1	Outer membrane protein YfgL, lipoprotein component of the protein assembly complex (forms a complex with YaeT, YfiO, and NlpB) CDS	1179	OuterMembrane	9.92	No internal helices
AIG69833.1	putative membrane protein CDS	1014	Unknown	2	1 internal helix
AIG69842.1	Thiosulfate sulfurtransferase, rhodanese CDS	846	Unknown	5.41	1 internal helix
AIG69880.1	putative alpha helix protein CDS	636	Unknown	2	No internal helices
AIG69913.1	hypothetical protein CDS	564	Unknown	2	No internal helices
AIG69916.1	hypothetical protein CDS	732	Unknown	2	No internal helices
AIG69918.1	putative component of the lipoprotein assembly complex (forms a complex with YaeT, YfgL, and NlpB) CDS	738	OuterMembrane	10	No internal helices
AIG69931.1	Signal recognition particle, subunit Ffh SRP54 CDS	1362	Unknown	5.6	No internal helices
AIG69958.1	Hypothetical protein CDS	1854	Unknown	2	No internal helices
AIG69974.1	Pertactin precursor CDS	4308	Unknown	5.87	No internal helices
AIG70043.1	Coenzyme F420 hydrogenase maturation protease	471	Unknown		Unknown
AIG70069.1	Lipoprotein NlpD CDS	1140	OuterMembrane	9.93	No internal helices
AIG70199.1	Type III secretion bridge between inner and outermembrane lipoprotein (YscJ,HrcJ,EscJ, PscJ) CDS	735	OuterMembrane	9.92	1 internal helix
AIG70202.1	Type III secretion protein EprH CDS	735	Unknown	2	1 internal helix
AIG70217.1	Uncharacterized protein YgeP CDS	1026	Unknown	2	No internal helices
AIG70219.1	putative lipoprotein YgeR precursor CDS	756	OuterMembrane	9.93	1 internal helix
AIG70290.1	Uridine kinase family protein CDS	714	Unknown	2	No internal helices
AIG70298.1	Putative metalloprotease yggG	759	Unknown	2	Unknown
AIG70300.1	hypothetical protein CDS	732	Unknown	2.5	No internal helices
AIG70312.1	UPF0301 protein YqgE CDS	564	Unknown	2	No internal helices

AIG70323.1	Uncharacterized protein YggN CDS	720	Unknown	2.5	No internal helices
AIG70398.1	Modulator of drug activity B CDS	582	Unknown	2	No internal helices
AIG70405.1	Ferrichrome-iron receptor CDS	2142	OuterMembrane	10	No internal helices
AIG70411.1	Type I secretion outer membrane protein, TolC precursor CDS	1482	OuterMembrane	10	No internal helices
AIG70414.1	Uncharacterized protein ygiD CDS	816	Unknown	2	No internal helices
AIG70516.1	type 1 fimbriae anchoring protein FimD CDS	2592	OuterMembrane	10	No internal helices
AIG70532.1	Putative lipid carrier protein CDS	525	Unknown	2	No internal helices
AIG70535.1	Putative protease	879	Unknown	2	Unknown
AIG70568.1	putative ABC transporter, auxiliary component YrbC CDS	636	Unknown	2.5	No internal helices
AIG70569.1	putative ABC transporter, periplasmic component YrbD CDS	552	Unknown	2.5	1 internal helix
AIG70575.1	Uncharacterized protein YrbK clustered with lipopolysaccharide transporters CDS	576	Unknown	2	1 internal helix
AIG70604.1	Outer membrane stress sensor protease DegQ, serine protease	1368	Periplasmic		Unknown
AIG70605.1	Outer membrane stress sensor protease DegS	1068	Periplasmic		Unknown
AIG70619.1	Rod shape-determining protein MreC CDS	1104	Unknown	2	1 internal helix
AIG70742.1	Type IV pilus biogenesis protein PilM CDS	795	Unknown	2	No internal helices
AIG70779.1	hypothetical protein CDS	993	Unknown	2.5	No internal helices
AIG70798.1	hypothetical protein CDS	912	Unknown	2	No internal helices
AIG70844.1	Putative transmembrane protein CDS	606	Unknown	2.5	No internal helices
AIG70880.1	hypothetical protein CDS	753	Unknown	2	No internal helices
AIG70893.1	TonB-dependent hemin , ferrichrome receptor CDS	1983	OuterMembrane	10	No internal helices
AIG70922.1	Uncharacterized protein YhjG CDS	2061	Unknown	2	2 internal helices
AIG70925.1	Protein YhjJ, putative peptidase CDS	1497	Unknown	2.5	No internal helices
AIG70935.1	hypothetical protein CDS	1560	Unknown	2	No internal helices
AIG70950.1	Putative fimbrial protein CDS	1056	Extracellular	9.65	1 internal helix
AIG70952.1	Long polar fimbria protein A precursor CDS	525	Extracellular	9.72	No internal helices
AIG70954.1	Putative lipase CDS	699	Unknown	7	1 internal helix
AIG70958.1	Outer membrane protein A precursor CDS	660	OuterMembrane	10	3 internal helices
AIG70960.1	putative exported protein CDS	711	Unknown	2.5	No internal helices
AIG70976.1	BAX protein CDS	573	Unknown	2	No internal helices
AIG70983.1	Putative outer membrane protein yiaT precursor CDS	741	OuterMembrane	10	1 internal helix
AIG70995.1	core protein CDS	4230	Unknown	2	No internal helices
AIG71007.1	hypothetical protein CDS	4767	OuterMembrane	9.95	2 internal helices
AIG71034.1	Lipopolysaccharide heptosyltransferase III CDS	1023	Unknown	2	No internal helices
AIG71060.1	Putative exported protein CDS	1710	Unknown	2	1 internal helix
AIG71083.1	Secreted protein EspB CDS	939	Extracellular	10	1 internal helix
AIG71084.1	Secreted protein EspD CDS	1125	Extracellular	10	3 internal helices
AIG71087.1	Type III secretion system EscD protein CDS	1221	Unknown	2	1 internal helix
AIG71088.1	Intimin CDS	2805	OuterMembrane	10	2 internal helices
AIG71090.1	translocated intimin receptor Tir CDS	1677	Extracellular	10	2 internal helices
AIG71096.1	SepQ CDS	918	Unknown	2	No internal helices
AIG71102.1	Type III secretion bridge between inner and outermembrane lipoprotein (YscJ,HrcJ,EscJ, PscJ) CDS	573	OuterMembrane	9.93	No internal helices

AIG71104.1	Type III secretion outermembrane pore forming protein (YscC,MxiD,HrcC, InvG) CDS	1539	OuterMembrane	10	No internal helices
AIG71117.1	ROrf2 CDS	1146	Extracellular	10	No internal helices
AIG71137.1	hypothetical protein CDS	624	Unknown	2	No internal helices
AIG71158.1	Uncharacterized protein YidR CDS	1251	Unknown	2	No internal helices
AIG71159.1	Uncharacterized protein YidS CDS	1083	Unknown	2	No internal helices
AIG71177.1	hypothetical protein CDS	750	Unknown	2	No internal helices
AIG71181.1	hypothetical protein CDS	2409	Unknown	5.48	No internal helices
AIG71182.1	hypothetical protein CDS	672	Unknown	2	No internal helices
AIG71192.1	Putative fimbrial protein CDS	1083	Extracellular	10	No internal helices
AIG71193.1	Putative fimbrial protein CDS	1071	Extracellular	9.65	No internal helices
AIG71194.1	type 1 fimbriae anchoring protein FimD CDS	2535	OuterMembrane	9.93	1 internal helix
AIG71272.1	hypothetical protein CDS	903	Unknown	2.5	No internal helices
AIG71273.1	hypothetical protein CDS	765	Unknown	2.5	No internal helices
AIG71277.1	hypothetical protein CDS	849	OuterMembrane	9.49	No internal helices
AIG71291.1	Putative carboxymethylenebutenolidase CDS	816	Unknown	2	No internal helices
AIG71333.1	Outer membrane sugar transport protein YshA CDS	693	OuterMembrane	9.93	1 internal helix
AIG71349.1	hypothetical protein CDS	915	Unknown	2	1 internal helix
AIG71360.1	hypothetical protein CDS	1056	Unknown	4.9	No internal helices
AIG71365.1	Putative glycoporin CDS	1395	Unknown	2.5	No internal helices
AIG71378.1	hypothetical protein CDS	675	Unknown	2	No internal helices
AIG71388.1	Putative uncharacterized protein YiiQ CDS	600	Unknown	2.5	1 internal helix
AIG71408.1	core protein CDS	4185	Unknown	2	No internal helices
AIG71411.1	hypothetical protein CDS	609	Unknown	2.5	No internal helices
AIG71419.1	hypothetical protein CDS	618	Unknown	2.5	No internal helices
AIG71443.1	Outer membrane vitamin B12 receptor BtuB CDS	1845	OuterMembrane	10	No internal helices
AIG71471.1	hypothetical protein CDS	636	Unknown	2	No internal helices
AIG71532.1	NMN phosphatase CDS	714	Unknown	2.5	1 internal helix
AIG71558.1	Putative exported protein CDS	690	Unknown	2.5	No internal helices
AIG71638.1	Outer membrane lipoprotein Blc CDS	534	OuterMembrane	10	1 internal helix
AIG71658.1	HflK protein CDS	1260	Unknown	5.48	1 internal helix
AIG71669.1	hypothetical protein CDS	639	Unknown	2	No internal helices
AIG71674.1	YjfP protein CDS	750	Unknown	2	No internal helices
AIG71687.1	Uncharacterized protein yjfZ CDS	795	Unknown	2	No internal helices
AIG71690.1	putative virulence protein CDS	1149	Unknown	2	No internal helices
AIG71691.1	Putative cell envelope opacity-associated protein A CDS	639	Unknown	2	No internal helices
AIG71701.1	Protein ytfJ precursor CDS	555	Unknown	4.9	No internal helices
AIG71707.1	Uncharacterized protein YtfM precursor CDS	1734	OuterMembrane	10	No internal helices
AIG71708.1	Uncharacterized protein YtfN CDS	3780	Unknown	4.72	1 internal helix
AIG71752.1	hypothetical protein CDS	1503	Unknown	2	No internal helices
AIG71763.1	hypothetical protein CDS	774	Unknown	2.5	No internal helices
AIG71781.1	hypothetical protein CDS	981	Unknown	2	No internal helices

AIG71788.1	type 1 fimbriae major subunit FimA CDS	549	Extracellular	10	1 internal helix
AIG71791.1	type 1 fimbriae anchoring protein FimD CDS	2637	OuterMembrane	10	No internal helices
AIG71794.1	mannose-specific adhesin FimH CDS	903	Unknown	4.65	2 internal helices
AIG71800.1	Uncharacterized protein YjiC CDS	831	Unknown	2	No internal helices
AIG71813.1	hypothetical protein CDS	843	Unknown	4.9	1 internal helix
AIG71857.1	hypothetical protein CDS	1074	Unknown	2	No internal helices
AIG68360.1	internalin,_putative_EDL933_2172	1260	Unknown	2	No internal helices

Table S2. Vaccine candidates expressed at transcriptional level (cutoff >10 RPKM)

Protein ID	Locus Tag	Functional annotation	Length (bp)	PSORT Localization	PSORT score	Expression (absolute value RPKM) RPKM (LB)	RPKM (LB agar)	RPKM (LB antibiotics)	RPKM (LB Feces)
AIG66267.1	EDL933_0059	hypothetical protein CDS	753	Unknown	2	32.7	65.6	75.0	0.0
AIG66347.1	EDL933_0139	Fimbrial protein Yad like protein CDS	1110	Unknown	2	192.8	327.0	101.7	347.5
AIG66348.1	EDL933_0140	Fimbrial protein YadK CDS	591	Unknown	2	32.2	37.6	0.0	171.7
AIG66349.1	EDL933_0141	Fimbrial protein YadL CDS	606	Unknown	2.5	44.4	48.9	0.0	134.0
AIG66424.1	EDL933_0216	putative aminopeptidase CDS	801	Unknown	2.5	9.8	18.5	105.7	0.0
AIG66447.1	EDL933_0241	hypothetical protein CDS	612	Unknown	2	223.3	250.2	288.3	779.5
AIG66490.1	EDL933_0284	putative tail fiber protein CDS	795	Unknown	2	12	0.0	1305.2	51.1
AIG66511.1	EDL933_0305	Zinc binding domain protein CDS	2334	Unknown	2	11	10.6	48.4	30.4
AIG66518.1	EDL933_0312	hypothetical protein CDS	1137	Unknown	2.5	24.6	0.0	24.8	0.0
AIG66656.1	EDL933_0452	Protein YkiA CDS	2166	Unknown	2	8.3	6.8	0.0	28.1
AIG66781.1	EDL933_0577	Putative stomatin/prohibitin-family membrane protease subunit YbbK	918	Unknown	2	45.2	225.9	0.0	619.2
AIG66972.1	EDL933_0770	core protein CDS	4200	Unknown	2	6.8	12.9	6.7	4.8
AIG66984.1	EDL933_0782	hypothetical protein CDS	1062	Unknown	2.5	21.6	18.6	13.3	38.2
AIG66987.1	EDL933_0786	putative fimbrial-like protein ygiL precursor CDS	567	Extracellular	9.72	41.5	17.4	24.9	143.2
AIG67020.1	EDL933_0819	TolA protein CDS	1185	Unknown	2	454.8	283.4	405.1	325.5
AIG67060.1	EDL933_0859	Putative outer membrane protein CDS	960	Unknown	2.5	45.5	30.9	338.2	126.9
AIG67245.1	EDL933_1046	Urease accessory protein UreF CDS	675	Unknown	2	23.2	0.0	0.0	0.0
AIG67308.1	EDL933_1109	putative hemolysin activator protein CDS	1065	OuterMembrane	8.86	0	0.0	13.3	19.1
AIG67453.1	EDL933_1258	hypothetical protein CDS	774	Unknown	2	68	51.0	383.0	131.1
AIG67464.1	EDL933_1259	Hypothetical protein CDS	1851	Unknown	2	24.9	27.4	78.4	0.0
AIG67503.1	EDL933_1308	hypothetical protein CDS	753	Extracellular	9.71	35.7	45.9	0.0	53.9
AIG67577.1	EDL933_1383	Shiga-like toxin II subunit A precursor CDS	960	Unknown	2.5	372.9	262.4	566.2	42.3
AIG67652.1	EDL933_1458	Putative exported protein precursor CDS	1086	Unknown	2	17.5	9.1	13.0	37.4
AIG67653.1	EDL933_1459	outer membrane fimbrial usher protein CDS	2523	OuterMembrane	10	11.5	7.8	0.0	72.4
AIG67671.1	EDL933_1477	hypothetical protein CDS	789	Unknown	2	2.8	0.0	0.0	25.7
AIG67672.1	EDL933_1478	hypothetical protein CDS	1389	Unknown	2	2.4	0.0	0.0	21.9
AIG67760.1	EDL933_1570	putative hemolysin activator protein CDS	1065	OuterMembrane	8.86	3.2	4.6	13.3	76.2
AIG67905.1	EDL933_1717	hypothetical protein CDS	588	Unknown	2	32.4	8.4	144.1	138.1
AIG67914.1	EDL933_1726	Hypothetical protein CDS	1938	Unknown	2	3.5	0.0	0.0	10.5
AIG67963.1	EDL933_1775	hypothetical protein CDS	744	Unknown	2	484.9	491.2	1537.0	2182.8
AIG68016.1	EDL933_1828	hypothetical protein CDS	1602	Unknown	2	3.8	0.0	8.8	12.7
AIG68165.1	EDL933_1977	Putative intestinal colonization factor encoded by prophage CP-933O CDS	747	Unknown	2	30	171.9	0.0	135.9
AIG68173.1	EDL933_1985	Hypothetical protein CDS	1854	Unknown	2	70.7	40.0	7.6	10.9
AIG68238.1	EDL933_2050	Hypothetical protein CDS	1851	Unknown	2	53.3	26.7	45.8	0.0
AIG68357.1	EDL933_2169	Outer membrane porin protein NmpC precursor CDS	1101	OuterMembrane	10	16.3	22.4	0.0	110.6
AIG68359.1	EDL933_2171	hypothetical protein CDS	669	Unknown	2	1.7	14.8	0.0	30.3
AIG68376.1	EDL933_2189	internalin, putative CDS	1260	Unknown	2	6.3	2.3	10.1	19.3
AIG68484.1	EDL933_2297	Hypothetical protein CDS	1854	Unknown	2	8.5	8.0	30.5	10.9
AIG68679.1	EDL933_2505	Hypothetical protein CDS	1851	Unknown	2	27.2	13.3	15.3	21.9
AIG68688.1	EDL933_2514	hypothetical protein CDS	585	Unknown	2	76.6	33.8	0.0	242.9
AIG69057.1	EDL933_2888	putative membrane protein CDS	651	Unknown	2	606.7	933.1	401.2	1262.9
AIG69153.1	EDL933_2986	hypothetical protein CDS	711	Unknown	2	108.7	83.4	119.1	142.8
AIG69189.1	EDL933_3022	Hypothetical protein CDS	1851	Unknown	2	18.8	10.7	30.5	0.0
AIG69198.1	EDL933_3031	hypothetical protein CDS	564	Unknown	2	37.7	78.8	25.0	180.0

AIG69346.1	EDL933_3180	Uncharacterized protein YehA precursor CDS	1035	Unknown	2	14.1	0.0	0.0	19.6
AIG69349.1	EDL933_3183	Putative fimbrial-like protein CDS	543	Unknown	2.5	88.7	582.1	208.0	37.4
AIG69390.1	EDL933_3224	Minor tail protein Z CDS	624	Unknown	2	3.6	0.0	22.6	0.0
AIG69393.1	EDL933_3227	Prophage Clp protease-like protein	1041	Unknown		2.2	0.0	13.6	0.0
AIG69411.1	EDL933_3245	Shiga toxin A-chain precursor CDS	948	Unknown	2.5	654.7	1586.3	871.2	235.5
AIG69664.1	EDL933_3500	Uncharacterized protein YadU in stf fimbrial cluster CDS	843	Unknown	2	33.2	93.7	50.2	72.2
AIG69958.1	EDL933_3799	Hypothetical protein CDS	1854	Unknown	2	31.4	16.0	22.8	21.9
AIG70298.1	EDL933_4145	Putative metalloprotease yggG	759	Unknown	2	519.6	390.4	93.0	320.9
AIG70798.1	EDL933_4653	hypothetical protein CDS	912	Unknown	2	3.7	0.0	15.5	22.3
AIG70844.1	EDL933_4699	Putative transmembrane protein CDS	606	Unknown	2.5	10.2	0.0	0.0	0.0
AIG71087.1	EDL933_4946	Type III secretion system EscD protein CDS	1221	Unknown	2	55.1	44.5	46.2	83.1
AIG71096.1	EDL933_4955	SepQ CDS	918	Unknown	2	223.3	102.2	399.8	309.6
AIG71137.1	EDL933_4996	hypothetical protein CDS	624	Unknown	2	19.8	39.6	22.6	130.1
AIG71181.1	EDL933_5041	hypothetical protein CDS	2409	Unknown	5.48	5.6	0.0	111.3	16.9
AIG71182.1	EDL933_5042	hypothetical protein CDS	672	Unknown	2	0	14.7	84.0	151.0
AIG71349.1	EDL933_5210	hypothetical protein CDS	915	Unknown	2	6.1	16.2	0.0	0.0
AIG71763.1	EDL933_5624	hypothetical protein CDS	774	Unknown	2.5	34.7	82.9	145.9	26.2
AIG71811.1	EDL933_5672	adherence and invasion outermembrane protein (Inv,enhances Peyer's patches colonization) CDS	5037	OuterMembrane	9.95	17.3	23.5	11.2	64.5
AIG66227.1	EDL933_0019	Putative outer membrane protein EDL933 0019	2451	OuterMembrane	10.00	3.2	10.1	0.0	8.3
AIG68053.1	EDL933_1865	Pertactin precursor	705	Unknown	7.00	36.6	42.0	40.0	28.8
AIG68360.1	EDL933_2172	internalin, putative	1260	Unknown	2.00	8	7.8	11.2	145.0
AIG69216.1	EDL933_3049	adherence and invasion outermembrane protein (Inv,enhances Peyer's patches colonization)	7863	OuterMembrane	9.95	5.8	12.6	10.8	10.3
AIG69974.1	EDL933_3815	Pertactin precursor	4308	Unknown	5.87	12	21.8	3.3	47.1
AIG68899.1	EDL933_2727	Putative lipoprotein	711	Unknown	2	11	6.9	0	0

Table S3. Features of the the 24 vaccine candidate expressed as recombinant proteins

No.	Protein ID	Vaccine candidate designation	Functional annotation	Recombinant protein	kDa
1	AIG67060.1	MC001	Putative outer membrane protein	I	35
2	AIG68165.1	MC002	Putative intestinal colonization factor encoded by prophage CP-933O	S	27
3	AIG68357.1	MC003	Outer membrane porin protein NmpC precursor	I	41
4	AIG67577.1	MC004	Shiga-like toxin II subunit A precursor	I	36
5	AIG69411.1	MC005	Shiga toxin A-chain precursor	S	35
6	AIG66347.1	MC006	Fimbrial protein Yad like protein	S	40
7	AIG66424.1	MC007	putative aminopeptidase	S	30
8	AIG66984.1	MC008	hypothetical protein	I	40
9	AIG67652.1	MC009	Putative exported protein precursor	I	38
10	AIG67671.1	MC010	hypothetical protein	I	31
11	AIG67672.1	MC011	hypothetical protein	I	51
12	AIG68053.1	MC012	Pertactin precursor	I	25
13	AIG68360.1	MC013	internalin, putative	S	48
14	AIG68899.1	MC014	putative lipoprotein	S	25
15	AIG69664.1	MC015	Uncharacterized protein YadU in stf fimbrial cluster	I	31
16	AIG70798.1	MC016	hypothetical protein	I	34
17	AIG67308.1	MC017	putative hemolysin activator protein 100%ID	I	37
18	AIG66972.1	MC018	core protein	S	159
19	AIG69216.1	MC019	adherence and invasion outermembrane protein (Inv,enhances Peyer's patches colonization)	S	275
20	AIG69974.1	MC020	Pertactin precursor	S	148
21	AIG71811.1	MC021	adherence and invasion outermembrane protein (Inv,enhances Peyer's patches colonization)	S	181
22	AIG71181.1	MC022	hypothetical protein	S	93
23	AIG66227.1	MC023	Putative outer membrane protein	S	92
24	AIG66656.1	MC024	Protein YkiA	I	83

S = soluble; I = insoluble

Subcellular localization (Psort, PDB)	Pfam domain	Expression (absolute value RPKM)				WB detection
		LB	LB Agar	LB Antibiotics	Feces	
OM	DUF2219	45.5	30.9	338.2	126.9	++
EC	SBP bac 11	30	171.9	0	135.9	++
OM	Porin 1	16.13	22.4	0	110.6	+
EC	RIP	372	262.4	566.2	42.3	+
EC	RIP	654.7	1586.3	871.2	235.5	-
OM	Fimbrial	192.8	327	101.7	347.5	-
Unknown	unknown	9.8	18.5	105.7	0	++
Fimbrial	Fimbrial	21.6	18.6	13.3	38.2	+
OM	Fimbrial	17.5	9.1	13	37.4	++
Unknown	DUF1329	2.8	0	0	25.7	+
OM	unknown	2.4	0	0	21.9	+
OM	unknown	36.6	42	40	28.8	+
Unknown	unknown	8	7.8	11.2	145	+
OM	SIMPL	11	6.9	0	0	-
Unknown	DUF2544	33.2	93.7	50.2	72.2	-
Unknown	DUF4225	3.7	0	15.5	22.3	-
OM	Potra 2	0	0	13.3	19.1	+
Unknown	unknown	6.8	12.9	6.7	4.8	+
OM	Invasin	5.8	12.6	10.8	10.3	-
OM	AIDA/Pertectin	12	21.8	3.3	47.1	++
OM	DUF3442	17.3	23.5	11.2	64.5	-
Unknown	unknown	5.6	0	111.3	16.9	+
OM	PapC Usher	3.2	10.1	0	8.3	+
Unknown	DUF2773	8.3	6.8	0	28.1	+

OM = Outermembrane; EC = extracellular

Chapter 2
Identification of the autochaperone domain in the Type Va secretion system (T5aSS): A prevalent feature of autotransporters with a β–helical passenger

Table 1S: Autochaperone (AC) domains identified following structural alignement among a 1523-entries dataset of well-defined and genuine autotransporters (ATs) (Celik et al. 2012).

ATs				AC prediction																																Functional genomics			Passenger fold prediction					
								BackPhyre														Phyre																						
								AC^{IcsA} query		AC^{EspP} query		AC^{Hbp} query		AC^{Pet} query		AC^{P69} query		AC^{Hap} query		AC^{IgA1} query		3ml3 template (AC I		3sze template (AC E		1wxr/3ak5 template		4om9 template (AC		1dab template (AC P		3syj template (AC Ha		3ho9 template (AC I										
Gene ID	Gene annotation	Species	AT sequence	AC sequence	Cluster	Confidence (%)	Distance TU-AC (a.a.)	Confidence (%)	Identity (%)	Confidence (%)	Identity (%)	Confidence (%)	Identity (%)	Confidence (%)	Identity (%)	Confidence (%)	Identity (%)	Confidence (%)	Identity (%)	Confidence (%)	Identity (%)	Confidence (%)	Identity (%)	Confidence (%)	Identity (%)	Confidence (%)	Identity (%)	Confidence (%)	Identity (%)	Confidence (%)	Identity (%)	Confidence (%)	Identity (%)	Confidence (%)	Identity (%)	Main category	Subcategory	Interpro	Beta-fold prediction	Passenger sequence	Phyre (PDB ; fold)	InterProScan (Beta-helix fold domain)	BetaWrap (Position)	PfScan Beta-solenoid (Position)
YP_0041	Hypothetical	Bartonella	MLQVFSSLL	SSISTMIL:	A	100.0	101	99.9	29	99.7	21	93.6	18	99.5	21	99.9	25	99.8	22	99.8	21	100.0	28	-	-	92.7	18	98.9	19	100.0	19	99.6	21	-	-	Autotransporter of unknc		IPR005546:	Yes	DTVYPKTK	1DAB;Cla	IPR011050	91-203	n.d.
YP_0041	Hypothetical	Bartonella	MLKVFKNYT	SLGLDNS	A	99.9	92	99.9	35	99.7	23	74.1	22	99.4	23	99.8	21	99.8	19	99.7	24	99.9	33	-	-	91.2	20	98.6	24	99.9	18	99.6	21	-	-	Autotransporter of unknc		IPR005546:	Yes	TSKVEDCS	1DAB;Cla	n.d.	91-203	L1 (71-132
YP_0016	Autotranspor	Bartonella	MIKVFRNHV	VSLMGGS	A	100.0	106	99.7	31	99.4	11	75.3	11	99.0	12	99.6	19	99.5	24	99.5	19	100.0	30	-	-	81.6	13	98.6	15	99.9	21	99.5	25	-	-	Autotransporter of unknc		IPR005546:	Yes	ATQEEGK	1DAB;Cla	IPR011050	428-554	n.d.
YP_0029	Autotranspor	Bartonella	MCACTTAIL	VSLMGES	A	100.0	96	99.7	28	99.2	13	-	-	98.6	13	99.6	21	99.4	22	99.3	20	100.0	32	-	-	94.4	14	98.9	15	100.0	21	99.6	24	-	-	Autotransporter of unknc		IPR005546:	Yes	ATEEVGKP	1DAB;Cla	n.d.	322-432	L1 (190-24
YP_0326	Hypothetical	Bartonella	MIKVFRNRV	SLVLLTNL	A	100.0	91	99.7	26	99.1	14	-	-	98.5	16	99.6	25	99.4	20	99.3	23	100.0	31	-	-	91.1	20	98.8	17	100.0	23	99.6	21	-	-	Autotransporter of unknc		IPR005546:	Yes	GEPYNITVV	1DAB;Cla	n.d.	91-203	L1 (188-29
YP_0041	Hypothetical	Bartonella	MIKVFNNYV	LSISVVSL	A	100.0	112	99.6	27	99.4	19	-	-	98.9	18	99.6	21	99.5	22	99.4	26	100.0	26	-	-	96.0	14	99.0	17	99.9	18	99.7	24	-	-	Autotransporter of unknc		IPR005546:	Yes	DQSSQCDI	1DAB;Cla	n.d.	91-203	L1 (96-188
YP_0323	Hypothetical	Bartonella	MNVEARSP	SSISFVRL	A	100.0	161	99.8	26	99.4	19	57.4	16	99.0	17	99.7	23	99.6	17	99.5	22	100.0	25	-	-	95.3	16	98.9	20	100.0	19	99.6	18	-	-	Autotransporter of unknc		IPR005546:	Yes	ALRAESGT	1DAB;Cla	n.d.	91-203	L1 (182-27
YP_0334	Hypothetical	Bartonella	MFKNCLYS	SSISFVKL	A	100.0	141	99.8	22	99.5	20	31.7	21	99.1	16	99.7	24	99.6	17	99.6	22	100.0	23	-	-	93.4	18	98.8	21	99.9	21	99.7	18	-	-	Autotransporter of unknc		IPR005546:	Yes	IIQANRVQI	1DAB;Cla	IPR011050	91-203	L1 (89-167
YP_0029	Autotranspor	Bartonella	MITKVSKNH	SFISFVKI	A	100.0	86	99.8	22	99.5	20	72.9	18	99.2	17	99.8	22	99.7	17	99.7	20	100.0	23	-	-	96.2	19	99.0	19	99.9	19	99.7	17	-	-	Autotransporter of unknc		IPR005546:	Yes	HAFVSLSC	1DAB;Cla	n.d.	473-603	L1 (111-17
YP_0016	Autotranspor	Bartonella	MVIKISNNR	SFISFVKL	A	100.0	85	99.8	23	99.5	21	22.7	19	99.1	15	99.7	20	99.6	18	99.6	19	100.0	23	-	-	95.5	23	98.8	19	99.9	17	99.7	17	-	-	Autotransporter of unknc		IPR005546:	Yes	RSFVSLSC	1DAB;Cla	n.d.	473-603	L1 (206-29
YP_0029	Autotranspor	Bartonella	MVINVFYKH	SSLSSVSI	A	100.0	359	99.6	28	99.3	22	-	-	98.8	19	99.5	24	99.5	23	99.3	16	100.0	25	-	-	95.4	19	98.9	22	100.0	20	99.7	24	-	-	Autotransporter of unknc		IPR005546:	Yes	AGSLCDPE	1DAB;Cla	IPR011050	99-202	L1 (70-135
YP_0029	Autotranspor	Bartonella	MVYVLKRH	SSLSSLSL	A	99.9	292	98.4	29	98.1	24	-	-	96.7	18	98.6	24	98.2	27	97.9	17	99.9	24	-	-	32.1	23	98.5	18	99.9	24	99.4	26	-	-	Autotransporter of unknc		IPR005546:	Yes	AGSSSDAC	1DAB;Cla	n.d.	322-432	n.d.
YP_0016	Autotranspor	Bartonella	MVYLFHKH	SSLSSLHI	A	99.9	409	99.1	25	98.6	20	-	-	97.9	16	99.1	27	98.9	25	98.8	16	99.9	23	-	-	-	-	98.2	14	99.9	24	99.2	21	-	-	Autotransporter of unknc		IPR005546:	Yes	DGTVNFNC	1DAB;Cla	n.d.	428-554	n.d.
YP_0016	Autotranspor	Bartonella	MVNVLRGY	SSLSGME	A	100.0	229	98.8	16	98.2	21	-	-	97.2	20	98.7	24	98.5	25	98.2	24	100.0	21	-	-	89.7	21	98.8	23	99.9	23	99.5	22	-	-	Autotransporter of unknc		IPR005546:	Yes	NSNNFATF	1DAB;Cla	IPR011050	428-554	L1 (76-193
YP_0016	Autotranspor	Bartonella	MFKNHLSLV	SSLSVVEI	A	100.0	295	98.5	28	97.8	19	-	-	96.1	18	98.4	22	98.2	19	97.8	17	100.0	27	-	-	95.9	22	99.0	18	100.0	16	99.7	18	-	-	Autotransporter of unknc		IPR005546:	Yes	GEVFSQKK	1DAB;Cla	IPR011050	428-554	L1 (193-27
YP_0029	Autotranspor	Bartonella	MLKNHLSLC	SFISSVNL	A	100.0	309	98.4	27	97.7	14	-	-	96.3	18	98.4	21	98.2	17	97.9	14	100.0	28	-	-	95.9	17	98.9	16	100.0	15	99.7	18	-	-	Autotransporter of unknc		IPR005546:	Yes	GEIPAEKKI	1DAB;Cla	IPR011050	322-432	L1 (193-27
YP_0029	Autotranspor	Bartonella	MVNLFKNRI	VDLSNSS	A	99.9	146	99.6	31	99.3	17	-	-	98.9	16	99.5	17	99.5	18	99.4	16	99.9	30	-	-	94.9	13	98.4	18	99.9	13	99.5	19	-	-	Autotransporter of unknc		IPR005546:	Yes	HNAESQVF	1DAB;Cla	IPR011050	322-432	L1 (115-17
YP_0016	Putative autc	Bartonella	MVNLLKNRI	SSLIHLV	A	100.0	90	99.7	28	99.4	17	33.5	17	99.0	18	99.6	26	99.6	24	99.5	21	99.9	27	-	-	95.7	16	98.8	19	100.0	21	99.7	23	-	-	Autotransporter of unknc		IPR005546:	Yes	KVQDGAIG	1DAB;Cla	n.d.	428-554	L1 (126-19
YP_0016	Autotranspor	Bartonella	MASWFSFIF	VSSVRLIN	A	100.0	70	99.9	32	99.7	17	81.4	15	99.4	16	99.8	27	99.8	24	99.7	22	100.0	28	-	-	95.4	15	99.0	18	100.0	23	99.7	19	-	-	Autotransporter of unknc		IPR005546:	Yes	NPFSTRTD	1DAB;Cla	n.d.	428-554	n.d.
YP_0029	Autotranspor	Bartonella	MTNVFKKR	SCVSSVR	A	100.0	80	99.9	31	99.8	15	93.4	16	99.6	11	99.9	25	99.9	21	99.8	21	100.0	25	-	-	93.5	15	99.0	13	100.0	27	99.6	22	-	-	Autotransporter of unknc		IPR005546:	Yes	GWFSISLPI	1DAB;Cla	n.d.	322-432	L1 (239-31
YP_0340	Surface prote	Bartonella	MINVLKNRT	SCVSSIRL	A	100.0	111	99.9	28	99.9	14	95.4	16	99.7	16	99.9	23	99.9	21	99.9	20	100.0	23	-	-	95.1	15	99.0	15	100.0	24	99.7	23	-	-	Autotransporter of unknc		IPR005546:	Yes	GEGFPSYF	1DAB;Cla	n.d.	91-203	L1 (1184-1
YP_0326	Surface prote	Bartonella	MINVFKNRT	SSISALHL	A	100.0	88	99.9	28	99.8	16	92.9	15	99.6	16	99.9	23	99.8	23	99.8	20	100.0	28	-	-	95.8	17	99.1	17	100.0	21	99.7	23	-	-	Autotransporter of unknc		IPR005546:	Yes	DTSPARSS	1DAB;Cla	n.d.	91-203	L1 (1046-1
YP_0029	Inducible bar	Bartonella	MYKKSLLSY	SRSDVST	A	100.0	72	99.1	22	98.5	19	-	-	97.8	19	98.9	28	98.9	25	98.6	26	100.0	19	-	-	93.6	14	99.0	17	100.0	24	99.7	24	-	-	Autotransporter of unknc		IPR005546:	Yes	ESLEASGE	1DAB;Cla	n.d.	473-603	L1 (165-25
YP_0016	Inducible aut	Bartonella	MYKKSLLLC	SDISVLSL	A	100.0	83	99.6	21	99.1	19	-	-	98.6	19	99.4	25	99.3	20	99.2	20	100.0	21	-	-	86.8	17	98.9	19	100.0	21	99.6	20	-	-	Autotransporter of unknc		IPR005546:	Yes	ENLEVADG	1DAB;Cla	n.d.	428-554	L1 (72-168
YP_0029	Inducible bar	Bartonella	MYKKSLLLY	SRSDISTL	A	100.0	74	99.5	23	99.2	19	-	-	98.9	17	99.3	27	99.3	22	99.0	17	100.0	22	-	-	89.5	16	98.9	15	100.0	22	99.7	22	-	-	Autotransporter of unknc		IPR005546:	Yes	ANNVLKND	1DAB;Cla	n.d.	473-603	L1 (249-32
YP_0029	Autotranspor	Bartonella	MYKTSLLSC	SDISTLHL	A	100.0	84	99.2	21	98.5	18	-	-	98.0	16	98.9	26	98.9	20	98.6	15	100.0	22	-	-	85.3	16	98.9	16	100.0	24	99.6	21	-	-	Autotransporter of unknc		IPR005546:	Yes	ETLEVSDG	1DAB;Cla	n.d.	473-603	L1 (212-28
YP_0016	Putative autc	Bartonella	MYKKFLLSY	SDVSVLN	A	100.0	84	99.7	23	99.3	23	-	-	98.9	19	99.5	28	99.3	21	99.2	17	100.0	21	-	-	87.8	16	98.9	19	100.0	27	99.7	23	-	-	Autotransporter of unknc		IPR005546:	Yes	KDLEVSDG	1DAB;Cla	n.d.	428-554	L1 (348-40
YP_0041	Autotranspor	Bartonella	MCKNYLLS	SSVSKLDI	A	99.9	69	99.2	23	98.1	21	-	-	97.5	20	98.8	22	98.4	22	98.5	28	99.9	26	-	-	-	-	98.4	30	99.9	30	99.3	30	-	-	Autotransporter of unknc		IPR005546:	Yes	EHKAEFEG	1DAB;Cla	n.d.	91-203	L1 (163-26
YP_0041	Hypothetical	Bartonella	MCKNYLSYS	STVSKLTL	A	99.9	64	99.3	23	98.6	19	-	-	97.9	19	99.1	23	99.0	25	98.9	25	99.9	27	-	-	52.8	25	98.8	23	99.9	26	99.4	28	-	-	Autotransporter of unknc		IPR005546:	Yes	NTEINSVYV	1DAB;Cla	n.d.	91-203	L1 (333-4
YP_9893	Outer membr	Bartonella	MNKKALLLC	SQISELKL	A	99.9	69	97.9	21	96.1	16	-	-	94.6	11	97.5	19	97.2	18	96.9	21	99.9	28	-	-	-	-	98.3	18	99.9	26	99.3	23	-	-	Autotransporter of unknc		IPR005546:	Yes	KTLVVEDG	1DAB;Cla	n.d.	352-483	L1 (55-124
YP_0016	Inducible aut	Bartonella	MCKKYIYKK	STVSVLNI	A	99.9	123	99.2	25	98.1	24	-	-	96.2	15	98.9	25	98.7	28	98.5	32	99.9	25	-	-	24.9	23	98.4	23	99.9	31	99.5	29	-	-	Autotransporter of unknc		IPR005546:	Yes	NTQPPEIID	1DAB;Cla	IPR011050	428-554	L1 (62-121
YP_0029	Inducible bar	Bartonella	MCKKYIYKK	STVSVLNI	A	99.9	123	99.2	22	98.0	19	-	-	97.3	12	98.8	28	98.8	21	98.5	30	99.9	24	-	-	38.1	25	98.5	27	99.9	26	99.5	30	-	-	Autotransporter of unknc		IPR005546:	Yes	NRQPPEIIS	1DAB;Cla	IPR011050	473-603	n.d.
YP_0340	Surface prote	Bartonella	MCKKSNKKI	LSTVSVLN	A	99.9	123	99.4	21	98.8	25	-	-	98.2	24	99.2	20	99.2	26	99.0	28	99.9	24	-	-	14.7	24	98.3	22	99.9	28	99.3	26	-	-	Autotransporter of unknc		IPR005546:	Yes	KTLLHHEIP	1DAB;Cla	IPR011050	91-203	n.d.
YP_0016	Inducible aut	Bartonella	MCKKYIYKK	SVVSVLNI	A	99.9	99	99.4	19	99.1	28	-	-	98.3	24	99.4	29	99.3	24	99.2	29	99.9	24	-	-	-	-	98.3	22	99.9	27	99.3	23	-	-	Autotransporter of unknc		IPR005546:	Yes	NTQPPEIP	1DAB;Cla	IPR011050	428-554	n.d.
YP_0029	Inducible bar	Bartonella	MCKKYIYKK	SIVSVLNL	A	99.9	103	99.6	21	99.3	21	-	-	98.5	22	99.6	23	99.5	25	99.4	28	99.9	25	-	-	9.6	19	98.3	24	99.9	24	99.3	23	-	-	Autotransporter of unknc		IPR005546:	Yes	NKQPLEIP	1DAB;Cla	IPR011050	473-603	n.d.
YP_0340	Hypothetical	Bartonella	MCKKSNKKI	SIVSVLNL	A	99.9	72	99.3	18	98.3	23	-	-	97.2	15	99.0	22	98.8	24	98.5	25	99.9	25	-	-	16.5	20	98.3	22	99.9	28	99.3	24	-	-	Autotransporter of unknc		IPR005546:	Yes	KTLLHHEIP	1DAB;Cla	IPR011050	91-203	n.d.
YP_0326	Hypothetical	Bartonella	MYKRNFLLC	SSVSVLNI	A	99.9	68	99.3	24	98.8	26	-	-	97.9	25	99.3	33	99.1	26	99.0	30	99.9	27	-	-	41.5	23	98.5	26	99.9	31	99.4	26	-	-	Autotransporter of unknc		IPR005546:	Yes	NTPPREIPI	1DAB;Cla	IPR011050	91-203	n.d.
YP_0029	Inducible bar	Bartonella	MKKSFLLYT	YSNLSGL	A	99.9	62	99.8	21	99.6	17	91.3	24	99.3	19	99.8	25	99.8	20	99.7	22	99.9	24	-	-	-	-	98.3	19	99.9	27	99.3	19	-	-	Autotransporter of unknc		IPR005546:	Yes	YALHDTNA	1DAB;Cla	IPR011050	473-603	L1 (279-34
YP_0016	Inducible aut	Bartonella	MKKSFLLYT	SFSNLSVI	A	99.9	82	99.8	23	99.5	18	-	-	99.0	17	99.7	25	99.6	18	99.6	24	99.9	23	-	-	-	-	98.4	18	99.9	23	99.2	18	-	-	Autotransporter of unknc		IPR005546:	Yes	ECPTNTSG	1DAB;Cla	n.d.	428-554	L1 (352-41
YP_0326	Hypothetical	Bartonella	MSKKSLLSY	NNSTIVFK	A	99.9	69	99.5	24	98.9	24	-	-	98.4	21	99.4	33	99.2	28	99.1	31	99.9	29	-	-	81.5	25	98.6	21	99.9	32	99.5	26	-	-	Autotransporter of unknc		IPR005546:	Yes	KSLSAGKG	1DAB;Cla	n.d.	91-203	L1 (191-26
YP_0340	Hypothetical	Bartonella	MSKKILLSY	HSEVFKLI	A	99.9	405	98.2	25	97.7	22	-	-	89.3	20	98.4	29	98.1	27	97.7	33	99.9	28	-	-	68.9	27	98.6	24	99.9	28	99.4	29	-	-	Autotransporter of unknc		IPR005546:	Yes	ISLFSGEGI	n.d.	n.d.	91-203	L1 (191-26
YP_0340	Hypothetical	Bartonella	MKKSFLLYT	SNLSGISL	A	99.9	66	99.6	24	99.3	22	-	-	98.8	22	99.5	24	99.5	21	99.3	24	99.9	29	-	-	16.3	20	98.6	24	99.9	22	99.4	26	-	-	Autotransporter of unknc		IPR005546:	Yes	STDLQPLF	1DAB;Cla	n.d.	91-203	n.d.
YP_0041	Inducible aut	Bartonella	MTRSLEAM	PESTVNV	A	99.9	213	98.8	16	98.1	28	-	-	96.6	19	98.8	23	98.4	21	98.1	18	99.9	25	-	-	35.9	23	98.7	23	99.9	23	99.6	20	-	-	Autotransporter of unknc		IPR005546:	Yes	QWIIPTDLI	1DAB;Cla	n.d.	91-203	n.d.
YP_0041	Autotranspor	Bartonella	MKKTFGIKL	STMHMLF	A	99.9	59	98.7	15	98.4	21	37.4	18	97.6	13	98.7	21	98.6	24	98.4	14	99.9	17	-	-	88.3	17	99.0	15	99.9	21	99.6	24	-	-	Autotransporter of unknc		IPR005546:	Yes	AELIADNGI	1DAB;Cla	IPR011050	91-203	L1 (77-158
YP_9893	Outer membr	Bartonella	MKDTAEGL	SSISFIHLI	A	99.9	239	98.7	25	98.2	16	-	-	97.6	17	98.8	17	98.4	21	98.1	19	99.9	29	-	-	87.9	19	99.0	19	99.9	16	99.4	21	-	-	Autotransporter of unknc		IPR005546:	Yes	SDRGSIELI	1DAB;Cla	n.d.	352-483	L1 (351-41
YP_9893	Outer membr	Bartonella	MLKKNILLC	SDLSSLK	A	100.0	77	99.9	24	99.7	23	83.5	24	99.3	22	99.8	27	99.8	21	99.7	25	100.0	21	-	-	83.3	18	99.0	23	99.9	28	99.5	21	-	-	Autotransporter of unknc		IPR005546:	Yes	SVPVDEIKF	1DAB;Cla	IPR011050	352-483	n.d.
YP_0016	Autotranspor	Bartonella	MIKVLKRHV	SSVSFVR	A	100.0	74	99.8	22	99.4	18	-	-	98.9	17	99.7	19	99.6	17	99.4	17	100.0	19	-	-	20.7	22	98.7	13	99.9	23	99.4	17	-	-	Autotransporter of unknc		IPR005546:	Yes	QEKSSCRS	1DAB;Cla	IPR011050	428-554	n.d.
YP_0024	Flagellin stru	Escherichi	MHSWKKKL	GVIAMNSI	Б	97.4	40	94.4	21.0	-	-	-	-	-	-	26.6	20	-	-	20.4	21	97.4	27	-	-	-	-	-	-	96.4	43	80.0	29	-	-	Autotransporter of unknc		IPR005546:	Yes	TTYNTFGY	1DAB;Cla	IPR011050;IPI	473-603	n.d.
YP_3093	Flagellar prot	Shigella sc	MHSWKKKL	GVIAMNSI	Б	97.6	41	85.9	21	23.5	9	-	-	-	-	24.5	20	-	-	-	-	97.6	27	-	-	-	-	-	-	96.5	43	86.2	29	-	-	Autotransporter of unknc		IPR005546:	Yes	ATYNTFGY	1DAB;Cla	IPR011050;IPI	180-297	n.d.
YP_2154	Flagellar prot	Salmonella	MHSWKKKL	GVVAMNS	Б	97.6	40	62.5	15	-	-	-	-	-	-	-	-	-	-	-	-	97.6	26	-	-	-	-	-	-	96.4	38	82.9	29	-	-	Autotransporter of unknc		IPR005546:	Yes	TTYDTFGY	n.d.	IPR011050;IPI	91-203	n.d.
YP_0021	Flagellin stru	Salmonella	MHSWKKKL	GVIAMTSI	Б	97.9	40	85.1	20	-	-	-	-	-	-	-	-	-	-	-	-	97.9	26	-	-	-	-	-	-	95.9	31	67.6	21	-	-	Autotransporter of unknc		IPR005546:	Yes	TTYNTFGY	1DAB;Cla	IPR011050;IPI	132-263	T8 (450-49
YP_0020	Putative autc	Salmonella	MWDDTAYV	GVIAMTSI	Б	97.0	138	22.3	15	-	-	-	-	-	-	-	-	-	-	-	-	97.0	28	-	-	-	-	-	-	94.8	33	68.8	24	-	-	Autotransporter of unknc		IPR005546:	Yes	LEGSGWF(	1DAB;Cla	IPR011050;IPI	132-263	T8 (337-38
NP_2861	Hypothetical	Escherichi	MHSWKKKL	GLIAMNSI	Б	97.6	44	95.9	19.0	45.0	7	-	-	-	-	38.3	13	22.5	15	43.7	20	97.6	26	-	-	-	-	-	-	95.8	33	86.3	22	-	-	Autotransporter of unknc		IPR005546:	Yes	ATNDISGQ	1DAB;Cla	IPR011050;IPI	n.d.	L1 (431-50
NP_3084	Flagellin stru	Escherichi	MHSWKKKL	GLIAMNSI	Б	97.8	39	97.8	15.0	64.3	12	-	-	43.2	7	86.5	14	75.5	14	85.7	21	97.3	27	-	-	-	-	-	-	96.3	43	82.3	21	-	-	Autotransporter of unknc		IPR005546:	Yes	ATNDISGQ	1DAB;Cla	IPR011050;IPI	n.d.	L1 (431-50
YP_0034	Flagellin stru	Escherichi	MHSWKKKL	GLIAMNSI	Б	97.3	39	96.5	15.0	22.8	8	-	-	-	-	44.0	14	-	-	44.1	19	97.3	27	-	-	-	-	-	-	95.6	33	83.5	21	-	-	Autotransporter of unknc		IPR005546:	Yes	ATNDISGQ	1DAB;Cla	IPR011050;IPI	265-383	n.d.
YP_0023	Flagellin stru	Escherichi	MHSWKKKL	GLIAMNSI	Б	97.2	42	97.2	19.0	58.8	11	-	-	-	-	-	-	-	-	24.2	10	97.2	23	-	-	-	-	-	-	94.8	31	81.9	20	-	-	Autotransporter of unknc		IPR005546:	Yes	ESNDISGT	2ZJ6;Clas	IPR011050;IPI	473-603	L1 (455-53
YP_0022	Autotranspor	Salmonella	MHSWKKKL	GVIAMNSI	Б	97.5	40	97.5	15.0	53.7	9	-	-	-	-	88.8	14	75.5	14	86.9	17	97.3	26	-	-	-	-	-	-	95.4	35	72.3	32	-	-	Autotransporter of unknc		IPR005546:	Yes	TTNDISGTT	1DAB;Cla	IPR011050;IPI	132-263	L1 (457-53
YP_0022	Autotranspor	Salmonella	MHSWKKKL	GVIAMNSI	Б	96.8	44	96.8	17.0	32.5	9	-	-	-	-	70.0	14	45.4	14	61.6	16	96.3	24	-	-	-	-	-	-	94.8	31	55.6	24	-	-	Autotransporter of unknc		IPR005546:	Yes	TTNDISGTT	1DAB;Cla	IPR011050;IPI	132-263	L1 (457-53
YP_0021	Autotranspor	Salmonella	MHSWKKKL	GVIAMNSI	Б	97.3	42	95.3	19.0	-	-	-	-	-	-	73.8	13	27.0	14	53.9	15	97.3	26	-	-	-	-	-	-	95.6	35	73.5	32	-	-	Autotransporter of unknc		IPR005546:	Yes	TTNDISGTT	1DAB;Cla	IPR011050;IPI	132-263	L1 (456-53
YP_0014	Hypothetical	Citrobacter	MHSWKKKL	GVIAMNSI	Б	97.7	44	95.8	20.0	-	-	-	-	-	-	76.9	15	60.0	19	65.5	18	97.7	27	-	-	-	-	-	-	96.5	33	91.1	22	-	-	Autotransporter of unknc		IPR005546:	Yes	STDISGTTY	1DAB;Cla	IPR011050;IPI	428-554	L1 (431-50
YP_0033	Outer membr	Citrobacter	MHSWKKKL	GIIAMNSD	Б	97.6	43	79.0	15	-	-	-	-	-	-	29.7	14	-	-	21.2	16	97.6	26	-	-	-	-	-	-	96.6	40	83.3	30	-	-	Autotransporter of unknc		IPR005546:	Yes	STDISGTTY	n.d.	IPR011050;IPI	265-383	L1 (431-50
YP_0023	Flagellin-like	Escherichi	MHSWKKKL	GVVAMNS	Б	97.5	40	66.9	19	-	-	-	-	-	-	-	-	-	-	-	-	97.5	26	-	-	-	-	-	-	95.8	33	77.1	27	-	-	Autotransporter of unknc		IPR005546:	Yes	ANYDTWG	n.d.	n.d.	473-603	n.d.
YP_5394	Structural prc	Escherichi	MHSWKKKL	GVVAMNS	Б	97.6	40	66.9	19	-	-	-	-	-	-	-	-	-	-	-	-	97.6	26	-	-	-	-	-	-	95.8	33	79.7	27	-	-	Autotransporter of unknc		IPR005546:	Yes	ANYDTWG	n.d.	n.d.	180-297	n.d.
YP_6683	Autotranspor	Escherichi	MHSWKKKL	GVVAMNS	Б	97.7	40	94.9	17.0	22.0	8	-	-	-	-	32.4	21	29.3	21	48.4	17	97.7	26	-	-	-	-	-	-	96.2	33	75.9	28	-	-	Autotransporter of unknc		IPR005546:	Yes	ANYDTWG	1DAB;Cla	n.d.	180-297	n.d.
YP_0032	Flagellin stru	Escherichi	MHSWKKKL	GIVAMNSI	Б	97.6	40	93.0	20.0	-	-	-	-	-	-	21.3	15	-	-	-	-	97.6	27	-	-	-	-	-	-	96.5	38	74.2	28	-	-	Autotransporter of unknc		IPR005546:	Yes	TTYDTWTY	1DAB;Cla	IPR011050;IPI	265-383	n.d.
YP_0030	Outer membr	Escherichi	MHSWKKKL	SSDYVLN.	Б	97.7	41	95.9	17.0	-	-	-	-	-	-	38.9	16	-	-	40.8	15	97.7	27	-	-	-	-	-	-	96.1	33	76.0	28	-	-	Autotransporter of unknc		IPR005546:	Yes	ANYDTWTY	1DAB;Cla	n.d.	322-432	L1 (425-50
YP_0014	Outer membr	Escherichi	MSGDSGGC	SSDYVLN.	Б	97.6	194	92.9	19.0	-	-	-	-	-	-	20.0	17	-	-	21.8	19	97.6	27	-	-	-	-	-	-	96.0	32	77.2	28	-	-	Autotransporter of unknc		IPR005546:	Yes	ETYTTESH	n.d.	n.d.	428-554	n.d.
YP_0024	Putative flag	Escherichi	MHSWKKKL	GIVAMNSI	Б	97.5	40	48.7	14	-	-	-	-	-	-	-	-	-	-	-	-	97.5	27	-	-	-	-	-	-	96.0	38	74.8	28	-	-	Autotransporter of unknc		IPR005546:	Yes	TTYDTWTY	1DAB;Cla	IPR011050;IPI	473-603	n.d.
NP_4593	Autotranspor	Salmonella	MHSWKKKL	GTIAMNSI	Б	97.5	40	79.4	18	-	-	-	-	-	-	25.6	14	-	-	-	-	97.5	26	-	-	-	-	-	-	96.5	39	80.6	30	-	-	Autotransporter of unknc		IPR005546:	Yes	TTYDTWTY	1DAB;Cla	IPR011050;IPI	n.d.	L1 (435-51
YP_0024	Putative flag	Escherichi	MHSWKKKL	GIVAMNSI	Б	97.5	40	35.4	19	-	-	-	-	-	-	-	-	-	-	-	-	97.5	27	-	-	-	-	-	-	96.4	40	74.6	28	-	-	Autotransporter of unknc		IPR005546:	Yes	TTYDTWTY	1DAB;Cla	IPR011050;IPI	473-603	n.d.
YP_0021	Autotranspor	Salmonella	MKNNVYFD	GTIAMNSI	Б	97.5	181	94.6	19	-	-	-	-	-	-	-	-	-	-	-	-	97.5	26	-	-	-	-	-	-	96.5	38	80.6	30	-	-	Autotransporter of unknc		IPR005546:	Yes	IGEDGQVA	1DAB;Cla	n.d.	132-263	n.d.
YP_0022	Outer membr	Salmonella	MGDVVFTSI	GVIAMTSI	Б	97.9	189	20.6	17	-	-	-	-	-	-	-	-	-	-	-	-	97.9	26	-	-	-	-	-	-	95.9	31	67.6	21	-	-	Autotransporter of unknc		IPR005546:	Yes	EVDTDHLT	n.d.	IPR011050;IPI	132-263	n.d.
YP_0032	Hypothetical	Cronobact	MQNGSEWI	LGQVDAD	Б	98.0	82	97.9	15.0	59.4	14	-	-	-	-	92.4	19	60.1	15	81.4	17	98.0	23	-	-	-	-	-	-	97.4	32	93.9	22	-	-	Autotransporter of unknc		IPR005546:	Yes	TLTEETASI	1DAB;Cla	IPR011050;IPI	265-383	n.d.
YP_0023	Virulence ass	Neisseria r	MAVDYVYDI	RWDSDV(	Б	100.0	106	100.0	25	99.9	19	99.2	22	99.9	14	99.9	21	99.9	21	99.9	18	95.8	27	-	-	-	-	7.2	15	30.6	19	11.6	21	-	-	Autotransporter of unknc		IPR005546:	Yes	ETEKPIRQ	1DAB;Cla	n.d.	473-603	n.d.
YP_9754	Virulence ass	Neisseria r	MAVDYVYDI	RWDSDV(	Б	100.0	108	100.0	26	99.9	20	99.2	24	99.9	15	100.0	22	100.0	22	100.0	19	95.8	28	-	-	-	-	11.1	17	34.2	19	8.4	22	-	-	Autotransporter of unknc		IPR005546:	Yes	EEEKTLQQ	1DAB;Cla	n.d.	352-483	n.d.
YP_2481	Autotranspor	Haemophil	MKLEASKQ/	RWNSNID	Б	100.0	77	100.0	23	99.9	19	99.2	21	99.9	16	100.0	23	100.0	21	99.9	20	95.0	26	-	-	-	-	13.3	19	-	-	9.1	21	-	-	Autotransporter of unknc		IPR005546:	Yes	VVDYTYDN	1DAB;Cla	n.d.	180-297	n.d.
YP_0012	Dna polymer	Haemophil	MAVDYVYDI	RWDSDVS	Б	99.9	129	99.9	21	99.8	18	97.6	19	99.6	16	99.8	19	99.8	20	99.8	21	54.5	17	-	-	-	-	-	-	-	-	-	-	-	-	Autotransporter of unknc		IPR005546:	Yes	DTEEQAKC	n.d.	n.d.	428-554	n.d.
YP_0041	Autotranspor	Haemophil	MIAFLFIPYS	RFNSDVH	Б	100.0	189	100.0	19	99.9	19	99.0	23	99.9	18	99.9	20	99.9	19	99.9	18	95.0	23	-	-	-	-	58.5	22	70.3	17	-	-	-	-	Autotransporter of unknc		IPR005546:	Yes	IDNYRSKFI	1DAB;Cla	IPR011050	91-203	n.d.
YP_0030	Aida-related	Neisseria r	MTLKASKQ/	RRHSDIHI	Б	99.9	63	99.9	21	99.7	17	98.3	13	99.6	16	99.9	17	99.9	20	99.8	20	93.5	22	-	-	-	-	8.4	18	6.4	15	-	-	-	-	Autotransporter of unknc		IPR005546:	Yes	SDAYDPVK	1DAB;Cla	n.d.	265-383	n.d.
YP_0040	Outer membr	Neisseria I	KQASKQASI	HTLETSDI	Б	99.9	60	99.9	21	99.7	21	98.7	13	99.7	16	99.9	17	99.9	18	99.9	17	95.1	24	-	-	-	-	11.8	20	5.7	16	-	-	-	-	Autotransporter of unknc		IPR005546:	Yes	SDAYDPVK	1DAB;Cla	n.d.	91-203	n.d.
YP_0023	Virulence ass	Neisseria r	MFRAQLGSI	RRHSDIHI	Б	100.0	166	100.0	20.0	99.7	17	98.0	13	99.6	14	99.9	16	99.9	15	99.9	17	92.3	23	-	-	-	-	13.1	19	-	-	-	-	-	-	Autotransporter of unknc		IPR005546:	Yes	TSLNNIFNK	1DAB;Cla	n.d.	473-603	n.d.
YP_0015	Virulence ass	Neisseria r	MFRAQLGSI	RRHSDIHI	Б	100.0	166	100.0	20.0	99.7	17	98.2	13	99.6	15	99.9	15	99.9	15	99.9	17	92.1	23	-	-	-	-	12.2	19	-	-	-	-	-	-	Autotransporter of unknc		IPR005546:	Yes	TSLNNIFNK	1DAB;Cla	n.d.	428-554	n.d.
YP_0022	Adhesin	Escherichi	MIASLFPAN	TVNGNLT	В	100.0	61	99.7	30	98.8	23	-	-	98.6	20	99.4	25	99.4	29	99.2	29	100.0	32	-	-	90.4	17	98.9	20	99.9	28	99.8	28	-	-	Autotransporter of unknc		IPR005546:	Yes	CQGYNIKA	1DAB;Cla	n.d.	473-603	L3 (350-44
YP_0030	Adhesin	Escherichi	MRVIFLRKE	TVNGNLT	В	100.0	73	99.8	30	98.8	22	-	-	98.6	19	99.5	25	99.6	29	99.4	28	100.0	33	-	-	93.9	16	98.7	20	99.9	28	99.8	29	-	-	Autotransporter of unknc		IPR005546:	Yes	AIDLCQGYI	1DAB;Cla	n.d.	265-383	L3 (366-46
YP_0035	Atp-binding c	Escherichi	MRVIFLRKE	TVNGNLT	В	100.0	73	99.8	33	98.9	21	-	-	98.8	19	99.6	26	99.6	29	99.5	24	100.0	33	-	-	94.2	15	99.0	19	100.0	26	99.8	28	-	-	Autotransporter of unknc		IPR005546:	Yes	AIDLCQGYI	1DAB;Cla	n.d.	200-322	L3 (366-46
YP_0024	Adhesin	Escherichi	MRIIFLRKEY	TVNGNLT	В	100.0	66	99.7	31	98.8	23	-	-	98.6	18	99.5	24	99.5	31	99.4	28	100.0	34	-	-	90.9	18	98.9	20	99.9	28	99.8	32	-	-	Autotransporter of unknc		IPR005546:	Yes	AAIDLCQG	1DAB;Cla	n.d.	473-603	L3 (366-46
YP_0022	Adhesin	Escherichi	MRIIFLRKEY	TVNGDLT	В	100.0	67	99.6	29	98.2	23	-	-	98.0	18	99.1	24	99.1	29	99.0	24	100.0	33	-	-	93.6	16	98.8	18	99.9	27	99.7	33	-	-	Autotransporter of unknc		IPR005546:	Yes	AAIDLCQG	1DAB;Cla	n.d.	473-603	L3 (366-46
YP_0032	Adhesin	Escherichi	MRIIFLRKEY	TVNGDLT	В	100.0	67	99.6	29	98.4	23	-	-	98.2	21	99.2	24	99.2	29	99.1	24	100.0	31	-	-	96.0	16	99.1	19	100.0	26	99.8	30	-	-	Autotransporter of unknc		IPR005546:	Yes	AAIDLCQG	1DAB;Cla	n.d.	265-383	L3 (366-46
YP_0023	Adhesin	Escherichi	MRIIFLRKEY	TVNGDLT	В	100.0	67	99.5	29	98.4	23	-	-	98.1	21	99.1	24	99.2	28	99.0	28	100.0	32	-	-	92.9	16	98.9	22	100.0	26	99.8	31	-	-	Autotransporter of unknc		IPR005546:	Yes	AAIDLCQG	1DAB;Cla	n.d.	473-603	L3 (366-46
YP_0032	Adhesin	Escherichi	MRIIFLRKEY	TVNGNLT	В	100.0	65	99.6	29	98.3	22	-	-	98.0	17	99.2	26	99.2	31	98.9	28	100.0	33	-	-	93.3	17	98.9	19	99.9	25	99.8	33	-	-	Autotransporter of unknc		IPR005546:	Yes	AAIDLCQG	1DAB;Cla	n.d.	265-383	L3 (366-46
YP_0032	Adhesin	Escherichi	MRIIFLRKEY	TVNGNLT	В	100.0	67	99.6	30	98.4	22	-	-	98.2	15	99.2	24	99.3	29	99.0	29	100.0	33	-	-	87.7	16	98.9	22	99.9	26	99.7	30	-	-	Autotransporter of unknc		IPR005546:	Yes	AAIDLCQG	1DAB;Cla	n.d.	265-383	L3 (366-46
YP_0014	Adhesin	Escherichi	MRIIFLRKEY	TVNGDLT	В	100.0	76	99.7	28	98.5	22	-	-	98.4	20	99.3	24	99.4	29	99.2	28	100.0	32	-	-	89.9	15	98.8	21	99.9	26	99.8	29	-	-	Autotransporter of unknc		IPR005546:	Yes	AIDLCQGYI	1DAB;Cla	n.d.	428-554	L3 (366-46
YP_0017	Adhesin	Escherichi	MRIIFLRKEY	TVNGNLT	В	100.0	72	99.6	31	98.5	21	-	-	98.4	17	99.3	25	99.3	27	99.2	29	100.0	32	-	-	69.7	15	98.4	20	99.9	28	99.7	29	-	-	Autotransporter of unknc		IPR005546:	Yes	AVTDSCQG	1DAB;Cla	n.d.	428-554	L3 (366-46
YP_0014	Adhesin	Escherichi	MRIIFLRKEY	TVNGNLT	В	100.0	73	99.6	29	98.1	22	-	-	98.0	17	99.2	25	99.3	29	98.9	29	100.0	33	-	-	86.9	16	98.7	19	99.9	27	99.7	29	-	-	Autotransporter of unknc		IPR005546:	Yes	AVTDSCQG	1DAB;Cla	n.d.	428-554	L3 (366-46
YP_0030	Adhesin	Escherichi	MIASLFSAN	TVNGNLT	В	100.0	66	99.5	28	98.3	23	-	-	98.0	17	99.0	22	99.1	28	98.9	28	100.0	33	-	-	91.5	16	98.8	18	99.9	26	99.7	27	-	-	Autotransporter of unknc		IPR005546:	Yes	CQGYDVK	1DAB;Cla	n.d.	322-432	L3 (350-44
NP_4167	Adhesin	Escherichi	MRIIFLRKEY	TVNGNLT	В	100.0	72	99.6	31	98.5	21	-	-	98.4	17	99.3	25	99.3	27	99.2	29	100.0	32	-	-	92.5	19	99.0	21	100.0	28	99.8	28	-	-	Autotransporter of unknc		IPR005546:	Yes	AVTDSCQG	1DAB;Cla	n.d.	n.d.	L3 (366-46
YP_0030	Adhesin	Escherichi	MRIIFLRKEY	TVNGNLT	В	100.0	72	99.7	28	98.6	22	-	-	98.4	19	99.3	25	99.4	28	99.2	22	100.0	30	-	-	93.7	15	98.8	16	99.9	25	99.8	27	-	-	Autotransporter of unknc		IPR005546:	Yes	AVTDSCQG	1DAB;Cla	n.d.	265-383	L3 (366-46
YP_0024	Adhesin	Escherichi	MRIIFLRKEY	TVNGNLT	В	100.0	73	99.5	28	98.0	23	-	-	97.8	18	99.0	24	99.1	29	98.8	24	100.0	33	-	-	94.3	15	99.0	17	100.0	25	99.8	28	-	-	Autotransporter of unknc		IPR005546:	Yes	AVIDSCQG	1DAB;Cla	n.d.	473-603	L3 (366-46
YP_0023	Adhesin	Escherichi	MRIIFLRKEY	TVNGDLT	В	100.0	75	99.8	34	99.0	23	-	-	98.8	20	99.6	25	99.6	29	99.5	29	100.0	32	-	-	91.2	16	98.9	21	99.9	23	99.7	29	-	-	Autotransporter of unknc		IPR005546:	Yes	AVIDSCQG	1DAB;Cla	n.d.	473-603	n.d.
YP_0017	Adhesin	Escherichi	MRIIFLRKEY	TVNGDLT	В	100.0	75	99.7	32	98.8	22	-	-	98.5	19	99.5	23	99.5	28	99.3	29	100.0	31	-	-	92.5	16	98.9	19	99.9	25	99.7	28	-	-	Autotransporter of unknc		IPR005546:	Yes	EVLDSCQG	1DAB;Cla	n.d.	428-554	n.d.
YP_0024	Adhesin	Escherichi	MRIIFLRKEY	TVNGNLT	В	100.0	75	99.7	32	98.7	23	-	-	98.2	16	99.4	23	99.4	29	99.2	27	100.0	32	-	-	92.3	17	98.8	20	99.9	25	99.8	29	-	-	Autotransporter of unknc		IPR005546:	Yes	EVLDSCQG	1DAB;Cla	n.d.	473-603	n.d.
YP_5415	Adhesin	Escherichi	MNMRIISLRI	SDVTVNG	В	100.0	75	99.6	29	98.2	21	-	-	98.0	17	99.2	23	99.2	29	99.0	23	100.0	32	-	-	93.8	16	98.9	20	99.9	26	99.8	31	-	-	Autotransporter of unknc		IPR005546:	Yes	AVIDSCQG	1DAB;Cla	n.d.	180-297	n.d.
YP_0023	Adhesin	Escherichi	MNMRIISLRI	TVNGNLT	В	100.0	79	99.6	29	98.2	21	-	-	98.0	17	99.2	23	99.2	29	99.0	23	100.0	31	-	-	84.6	15	98.7	21	99.9	26	99.7	31	-	-	Autotransporter of unknc		IPR005546:	Yes	AVIDSCQG	1DAB;Cla	n.d.	473-603	n.d.
YP_6701	Adhesin	Escherichi	MNMRIIFLRI	TVNGNLT	В	100.0	75	99.7	31	98.6	22	-	-	98.3	17	99.4	24	99.4	31	99.3	30	100.0	34	-	-	91.5	18	99.0	22	100.0	28	99.8	31	-	-	Autotransporter of unknc		IPR005546:	Yes	AVIDSCQG	1DAB;Cla	n.d.	180-297	n.d.
NP_7546	Adhesin	Escherichi	MNMRIISLRI	TVNGNLT	В	100.0	75	99.6	28	98.5	20	-	-	98.3	18	99.3	22	99.3	30	99.1	29	100.0	31	-	-	87.0	15	98.8	21	99.9	22	99.7	31	-	-	Autotransporter of unknc		IPR005546:	Yes	AVIDSCQG	1DAB;Cla	n.d.	n.d.	n.d.
YP_0023	Adhesin	Escherichi	MNMRIISLRI	TVNGNLT	В	100.0	73	99.7	31	98.7	21	-	-	98.5	18	99.5	22	99.5	30	99.3	28	100.0	32	-	-	91.5	15	98.7	17	99.9	21	99.7	30	-	-	Autotransporter of unknc		IPR005546:	Yes	AVIDSCQG	1DAB;Cla	n.d.	473-603	n.d.
YP_0023	Adhesin	Escherichi	MRIIFLRKEY	TVNGNLT	В	100.0	73	99.7	28	98.6	22	-	-	98.3	20	99.4	25	99.4	28	99.3	26	100.0	32	-	-	90.7	18	98.8	18	99.9	25	99.8	28	-	-	Autotransporter of unknc		IPR005546:	Yes	AVIDSCQG	1DAB;Cla	n.d.	473-603	n.d.
YP_0042	Outer membr	Rahnella s	MVFNKITAC	TVNGSLD	В	100.0	72	99.3	32	97.7	25	-	-	97.7	20	99.0	23	98.9	27	98.8	19	100.0	33	-	-	94.1	17	99.1	21	100.0	26	99.8	26	-	-	Autotransporter of unknc		IPR005546:	Yes	ETLPTCSG	1DAB;Cla	n.d.	91-203	L3 (361-45
YP_4039	Atpase	Shigella dy	MVDLAGGN	YLADVTVI	В	100.0	193	91.5	30	-	-	-	-	-	-	-	-	29.7	25	-	-	100.0	31	-	-	91.7	16	98.7	18	99.9	25	99.7	29	-	-	Autotransporter of unknc		IPR005546:	Yes	AVRVEEGT	1DAB;Cla	n.d.	180-297	n.d.
YP_8343	Outer membr	Burkholder	MNHSFRSIV	LNNAGLV.	Г	100.0	123	99.5	36	98.7	21	-	-	98.2	23	99.2	30	98.8	27	98.9	22	100.0	36	-	-	66.0	23	98.7	20	99.9	27	99.5	25	-	-	Autotransporter of unknc		IPR005546:	Yes	STCGNGAA	1DAB;Cla	n.d.	345-479	L1 (1060-1
YP_6200	Outer membr	Burkholder	MNHSFRSIV	LNNAGLV.	Г	100.0	123	99.5	36	98.7	21	-	-	98.2	23	99.2	30	98.8	27	98.9	22	100.0	36	-	-	80.5	20	98.6	18	99.9	28	99.5	24	-	-	Autotransporter of unknc		IPR005546:	Yes	STCGNGAA	1DAB;Cla	n.d.	180-297	L1 (1060-1
YP_0022	Putative oute	Burkholder	MNHSFRSIV	LTNAGLV/	Г	100.0	121	99.5	34	98.7	20	-	-	98.2	19	99.3	28	98.8	28	99.0	23	100.0	34	-	-	96.4	27	99.2	20	100.0	32	99.7	27	-	-	Autotransporter of unknc		IPR005546:	Yes	STCGNGAA	1DAB;Cla	n.d.	132-263	L1 (1059-1
YP_0017	Outer membr	Burkholder	MNHSFRSIV	LNNAGLV.	Г	100.0	123	99.7	38	99.3	21	-	-	98.9	21	99.5	30	99.3	25	99.3	23	100.0	37	-	-	97.5	26	99.3	23	100.0	32	99.8	27	-	-	Autotransporter of unknc		IPR005546:	Yes	STCGNGTA	1DAB;Cla	n.d.	428-554	L1 (1060-1

YP_3680	Outer membr	Burkholder	MNHSFRSIV	LNNAGLV.	Г	100.0	114	99.5	34	98.8	20	-	-	98.4	21	99.3	30	98.9	25	99.1	24	100.0	33	-	-	89.0	23	98.9	20	99.9	30	99.6	23	-	-	Autotransporter of unknc	IPR005546:	Yes	STCGGGT/	1DAB;Cla	n.d.	180-297	L1 (1128-
YP_7724	Outer membr	Burkholder	MRFHRTAVI	LNNAGLV.	Г	100.0	134	99.6	32	99.0	21	-	-	98.6	22	99.5	31	99.1	25	99.3	23	100.0	35	-	-	81.3	25	99.0	20	100.0	32	99.6	23	-	-	Autotransporter of unknc	IPR005546:	Yes	STCGDGS/	1DAB;Cla	n.d.	180-297	L1 (1146-
YP_0018	Outer membr	Burkholder	MRFHRTAVI	LNNAGLV.	Г	100.0	134	99.7	33	99.2	20	-	-	98.9	22	99.6	33	99.4	28	99.5	22	100.0	37	-	-	90.6	27	98.9	22	99.9	33	99.7	24	-	-	Autotransporter of unknc	IPR005546:	Yes	STCGDGS/	1DAB;Cla	n.d.	428-554	L1 (1146-
YP_0019	Outer membr	Burkholder	MNHSFRSIV	LTNAGTV.	Г	100.0	109	99.5	33	98.8	19	-	-	98.4	19	99.4	31	98.9	28	99.1	25	100.0	33	-	-	92.2	26	99.1	19	99.9	29	99.7	25	-	-	Autotransporter of unknc	IPR005546:	Yes	STCGDGTS	1DAB;Cla	n.d.	132-263	L1 (1130-
YP_0015	Outer membr	Burkholder	MTASLSPLF	TGDSAVC	Д	100.0	113	98.7	30	97.2	20	-	-	96.8	19	97.6	27	97.6	29	96.8	28	100.0	31	-	-	96.2	24	99.4	22	100.0	32	99.8	28	-	-	Autotransporter of unknc	IPR005546:	Yes	QVVNNPQ\	1DAB;Cla	n.d.	428-554	L1 (109-1
YP_0018	Outer membr	Burkholder	MNRRRFPF	NSSVGSV	Д	100.0	35	96.3	32	63.3	18	-	-	78.3	16	76.9	30	75.8	25	76.4	19	100.0	36	-	-	89.7	27	99.3	21	100.0	30	99.8	24	-	-	Autotransporter of unknc	IPR005546:	Yes	QVVSGPQ(	1DAB;Cla	n.d.	428-554	L1 (90-15
YP_0011	Outer membr	Burkholder	MNHRSFPL	DSSVGSV	Д	100.0	92	99.2	33	98.1	20	-	-	97.6	17	98.7	28	98.8	27	98.5	19	100.0	35	-	-	92.4	26	99.2	20	100.0	27	99.8	29	-	-	Autotransporter of unknc	IPR005546:	Yes	AGGPQQAI	1DAB;Cla	n.d.	428-554	L1 (90-15
NP_7531	Antigen 43 (	Escherichi	MQTHRHEI(	DLSHAGC	Е	100.0	119	99.7	31	98.9	23	-	-	98.4	22	99.5	22	99.3	24	99.4	24	100.0	30	-	-	21.9	22	98.7	21	99.9	20	99.4	22	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADSVVPAG	1DAB;Cla	n.d.	n.d.	T4 (195-2
YP_0024	Antigen 43 (	Escherichi	MLMKRHLN	DLSHAGC	Е	100.0	69	99.5	25	98.7	19	-	-	98.4	21	99.3	23	99.0	21	99.1	24	100.0	27	-	-	27.6	25	98.7	23	99.9	20	99.5	21	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADIVVHPGI	1DAB;Cla	n.d.	473-603	T4 (109-1
YP_0024	Antigen 43 (	Escherichi	MLMKRHLN	DLSHAGC	Е	100.0	69	99.3	26	98.3	18	-	-	97.7	24	99.0	24	98.6	22	98.7	24	100.0	27	-	-	90.1	25	99.0	23	100.0	23	99.7	20	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADIVVHPGI	1DAB;Cla	n.d.	473-603	T4 (109-1
NP_7555	Antigen 43 (	Escherichi	MLMKRHLN	DLSHAGC	Е	100.0	69	99.1	24	98.1	18	-	-	97.5	22	98.9	23	98.4	20	98.6	22	100.0	28	-	-	94.2	27	99.0	22	100.0	22	99.7	20	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADIVVHPGI	1DAB;Cla	n.d.	n.d.	T4 (145-2
YP_6682	Autotranspor	Escherichi	MLMKRHLN	DLSHAGC	Е	100.0	70	99.2	24	98.1	18	-	-	97.3	20	98.9	23	98.4	21	98.7	25	100.0	27	-	-	92.2	27	99.1	23	100.0	21	99.7	20	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADIVVHPGI	1DAB;Cla	n.d.	180-297	T4 (145-2
YP_0024	Antigen 43 (	Escherichi	MLMKRHLN	DLSHAGC	Е	100.0	69	99.0	25	98.0	19	-	-	97.1	21	98.6	23	98.2	23	98.3	24	100.0	30	-	-	24.3	23	98.5	21	99.9	20	99.5	24	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADIVVHPGI	1DAB;Cla	n.d.	473-603	L1 (425-5
YP_0024	Antigen 43 (	Escherichi	MLMKRHLNTS	DLSHAGC	Е	100.0	67	98.6	25	96.7	18	-	-	94.4	21	98.1	25	97.3	21	97.7	24	100.0	28	-	-	45.8	26	98.6	24	99.9	22	99.5	21	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADKVVQAC	1DAB;Cla	n.d.	473-603	T4 (107-1
AP_0025	Antigen 43 (	Escherichi	MLMKRHLNTC	DLSHAGC	Е	100.0	67	98.9	26	97.7	18	-	-	96.5	21	98.5	22	97.9	22	98.1	24	100.0	27	-	-	88.4	26	98.8	22	99.9	21	99.6	21	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADIVVHPGI	1DAB;Cla	n.d.	n.d.	T4 (107-1
NP_8384	Outer membr	Shigella fle	MEGKADNV	DLSHAGC	Е	100.0	71	88.3	23	-	-	-	-	-	-	23.3	21	-	-	-	-	100.0	27	-	-	39.1	26	98.7	23	99.9	20	99.5	21	-	-	Self-Associating AutoTra	IPR030930:	Yes	DSGAAVS(	1DAB;Cla	n.d.	n.d.	T5 (95-15
YP_0032	Aida-i adhesi	Escherichi	MKRHLNTS	AVLSRAV.	Ж	100.0	67	99.3	27	98.4	17	-	-	98.0	15	99.0	23	98.9	23	98.9	22	100.0	25	-	-	92.2	24	99.0	17	99.9	17	99.7	24	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADTVVQAC	1DAB;Cla	n.d.	265-383	T4 (107-1
YP_0032	Aida-i adhesi	Escherichi	MKRHLNTS	YFDNQTT	Ж	100.0	67	99.1	23	97.8	17	-	-	96.8	16	98.7	21	98.5	21	98.6	17	100.0	23	-	-	12.6	18	98.4	20	99.9	16	99.4	22	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADTVVQAC	1DAB;Cla	n.d.	265-383	T4 (107-1
YP_0033	Adhesin autc	Citrobacter	MKRHLNTS	AALSRAV.	Ж	100.0	67	99.1	22	97.6	20	-	-	97.0	18	98.6	24	98.2	21	98.3	17	100.0	23	-	-	-	-	98.1	16	99.9	17	99.4	21	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADTVVQAC	1DAB;Cla	n.d.	265-383	T4 (107-1
YP_0017	Antigen 43	Escherichi	MKRHLNTS	AVLSRAV.	Ж	100.0	67	99.1	24	97.7	20	-	-	97.2	16	98.7	23	98.4	23	98.6	21	100.0	23	-	-	90.8	24	98.7	19	99.9	19	99.6	25	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADTVVEAG	1DAB;Cla	n.d.	428-554	T4 (107-1
YP_0030	Antigen 43 (	Escherichi	MKRHLNTS	AALSRAV.	Ж	100.0	67	98.5	24	95.9	17	-	-	94.3	18	97.8	23	97.3	23	97.7	17	100.0	24	-	-	13.4	23	98.5	18	99.9	19	99.5	25	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADTVVQAC	1DAB;Cla	n.d.	265-383	T4 (107-1
YP_6708	Autotranspor	Escherichi	MKRHLNTS	AALSRAV.	Ж	100.0	67	99.4	27	98.4	20	-	-	98.0	16	99.1	23	98.9	24	99.0	22	100.0	25	-	-	73.0	26	98.8	16	99.9	19	99.6	24	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADTVVQAC	1DAB;Cla	n.d.	180-297	T4 (107-1
YP_0014	Antigen 43	Escherichi	MKRHLNTS	AVLSRAV.	Ж	100.0	67	99.5	23	98.7	17	-	-	98.4	17	99.1	22	99.0	21	99.1	20	100.0	23	-	-	17.8	22	98.3	18	99.9	19	99.5	21	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADTVVQAC	1DAB;Cla	n.d.	428-554	T4 (107-1
YP_0014	Antigen 43 (	Escherichi	MKRHLNTS	AALSRAV.	Ж	100.0	67	99.5	26	98.7	17	-	-	98.4	17	99.1	22	99.1	21	99.1	19	100.0	25	-	-	-	-	98.1	18	99.9	17	99.3	21	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADTVVQAC	1DAB;Cla	n.d.	428-554	T1 (425-4
YP_5401	Autotranspor	Escherichi	MKRHLNTS	AALSRAV.	Ж	99.9	67	99.3	23	98.1	19	-	-	97.6	16	98.9	23	98.6	18	98.8	17	99.9	24	-	-	8.1	20	97.8	18	99.9	17	99.4	21	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADTVVQAC	1DAB;Cla	n.d.	180-297	T4 (107-1
YP_0032	Aida-i adhesi	Escherichi	MQNGIAHN	AALSRAV.	Ж	100.0	123	99.5	25	98.6	20	-	-	98.3	17	99.2	22	99.1	24	99.0	22	100.0	25	-	-	21.5	27	98.7	16	99.9	17	99.5	25	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADKVVQAC	1DAB;Cla	n.d.	265-383	T4 (199-2
YP_0023	Antigen 43 p	Escherichi	MLFCTCRY	AALSRAV.	Ж	100.0	92	99.4	21	98.5	19	-	-	98.2	17	99.0	22	98.8	22	98.8	24	100.0	24	-	-	10.1	25	98.5	18	99.9	15	99.5	23	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADTVVQAC	1DAB;Cla	n.d.	473-603	T4 (132-2
NP_3094	Aida-i	Escherichi	MKRHLNTS	AALSRAV.	Ж	100.0	67	99.3	25	98.3	20	-	-	97.8	16	98.9	19	98.6	25	98.7	25	100.0	25	-	-	32.5	27	98.6	18	99.9	16	99.4	22	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADKVVQAC	1DAB;Cla	n.d.	n.d.	T4 (143-2
YP_0029	Outer membr	Variovorax	MNKTFRTL	ANSGIVSI	З	100.0	83	95.7	38	-	-	-	-	-	-	55.7	28	29.6	28	20.1	22	100.0	41	-	-	-	-	98.7	15	99.9	28	99.6	26	-	-	Autotransporter of unknc	IPR005546:	Yes	QGGAGGQ	1DAB;Cla	n.d.	473-603	TS (486-5
YP_0041	Outer membr	Variovorax	MNKIFRTL	ANSGTVS	З	100.0	94	97.9	39	84.5	15	-	-	85.5	11	96.1	25	95.2	34	91.7	21	100.0	41	-	-	-	-	98.7	15	99.9	25	99.7	33	-	-	Autotransporter of unknc	IPR005546:	Yes	AGGAGGQ	n.d.	IPR011050;IPF	91-203	n.d.
YP_0029	Outer membr	Variovorax	MNNIHCSV	ANSGTVS	З	100.0	57	98.0	43	87.8	18	-	-	85.9	10	94.7	30	96.4	31	93.1	33	100.0	40	-	-	13.6	18	98.9	17	99.9	28	99.7	31	-	-	Autotransporter of unknc	IPR005546:	Yes	AGGGAAM	n.d.	n.d.	473-603	L1 (275-3
YP_5827	Hypothetical	Cupriavidu	MNHIYRLV	AGLNNGC	З	100.0	111	99.4	39	97.9	22	-	-	97.6	21	98.7	36	98.8	28	98.3	25	100.0	35	-	-	28.4	15	99.0	23	100.0	30	99.6	29	-	-	Autotransporter of unknc	IPR005546:	Yes	ATPIAGTLF	1DAB;Cla	n.d.	180-297	L3 (304-3
YP_0041	Outer membr	Variovorax	MNTIFRVIW	ANSGTVS	З	100.0	97	96.6	32	-	-	-	-	28.6	11	84.3	28	64.4	23	46.5	27	100.0	36	-	-	53.3	19	99.0	17	99.9	27	99.7	31	-	-	Autotransporter of unknc	IPR005546:	Yes	ACTTDTAA	n.d.	n.d.	91-203	L3 (451-5
YP_0039	Outer membr	Achromob	MNHVYRVV	ASLNNSG	З	100.0	85	99.9	42	99.4	17	-	-	99.2	15	99.7	30	99.7	28	99.5	21	100.0	43	-	-	77.8	16	98.8	13	99.9	29	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	AGGNPPLH	1DAB;Cla	n.d.	91-203	L3 (141-2
YP_0041	Outer membr	Variovorax	MNTAYKTI	AGLANSG	З	100.0	96	98.6	36	90.6	21	-	-	87.3	18	97.1	28	97.1	29	95.5	22	100.0	39	-	-	37.3	23	98.8	17	99.9	27	99.6	29	-	-	Autotransporter of unknc	IPR005546:	Yes	ACAPAAPA	1DAB;Cla	IPR011050;IPF	91-203	L3 (152-2
YP_0041	Outer membr	Variovorax	MTSKRMNR	PSLANSG	З	100.0	56	98.8	35	96.5	15	-	-	94.4	16	97.7	36	98.0	25	96.5	23	100.0	36	-	-	11.5	18	98.9	17	99.9	31	99.6	26	-	-	Protease / Peptidase S6	IPR005546:	Yes	IQTYAGDP	1R6V;Cla	n.d.	91-203	n.d.
YP_0022	Putative autc	Burkholder	MKRKARNL	LTNAGRV	И	100.0	93	73.9	35	-	-	-	-	-	-	-	-	-	-	-	-	100.0	39	-	-	81.8	28	98.7	25	99.9	34	99.7	33	-	-	Autotransporter of unknc	IPR005546:	Yes	ACSATAPG	1DAB;Cla	n.d.	132-263	L3 (532-6
YP_3736	Outer membr	Burkholder	MKRTTRNL	LTNAGLV	И	100.0	93	94.6	27	-	-	-	-	-	-	-	-	-	-	-	-	100.0	31	-	-	16.1	18	98.5	14	99.9	29	99.7	27	-	-	Autotransporter of unknc	IPR005546:	Yes	ACIPAAPP	n.d.	n.d.	180-297	TS (443-4
YP_0017	Outer membr	Burkholder	MAILTGAVP	LTNAGLV	И	100.0	81	92.7	37	-	-	-	-	-	-	-	-	33.3	24	-	-	100.0	39	-	-	80.2	28	98.7	23	99.9	35	99.7	30	-	-	Autotransporter of unknc	IPR005546:	Yes	ACSATAPG	1DAB;Cla	n.d.	428-554	L3 (493-5
YP_0016	Outer membr	Pseudomo	MRCHRLLL	NAGTIDLS	Й	100.0	116	98.3	30	94.7	17	-	-	90.1	19	97.4	23	96.5	29	94.1	22	100.0	31	-	-	13.0	17	98.8	16	100.0	25	99.6	28	-	-	Autotransporter of unknc	IPR005546:	Yes	GENLGALV	1DAB;Cla	n.d.	428-554	n.d.
YP_6087	Outer membr	Pseudomo	MRYLPLLAF	NAGLIDLS	Й	100.0	78	99.0	34	97.4	17	-	-	96.6	18	98.4	27	98.1	29	97.5	24	100.0	35	-	-	74.2	20	98.9	16	100.0	27	99.7	29	-	-	Autotransporter of unknc	IPR005546:	Yes	QTLTVSPG	1DAB;Cla	IPR011050	180-297	n.d.
NP_7452	Outer membr	Pseudomo	MEVFLMRLF	NAGTLDL	Й	100.0	96	98.2	32	88.0	20	-	-	89.5	13	97.7	23	96.2	27	94.2	19	100.0	33	-	-	52.6	15	99.0	16	100.0	27	99.7	25	-	-	Autotransporter of unknc	IPR005546:	Yes	ACTFSPGP	1DAB;Cla	n.d.	n.d.	n.d.
NP_8828	Autotranspor	Bordetella	MRRLKAQA	GDVANRC	К	100.0	72	99.8	36	98.8	17	-	-	98.8	17	99.4	20	99.2	27	99.2	18	100.0	41	-	-	56.9	23	98.7	17	99.9	21	99.5	25	-	-	Autotransporter of unknc	IPR005546:	Yes	AWANCTTS	1DAB;Cla	n.d.	n.d.	L3 (610-6
NP_8793	Autotranspor	Bordetella	MGSGPSGN	GDVANRC	К	100.0	128	99.9	40	99.2	17	-	-	99.1	17	99.6	24	99.5	27	99.4	22	100.0	42	-	-	56.5	23	98.7	18	99.9	22	99.5	27	-	-	Autotransporter of unknc	IPR005546:	Yes	AGNTVTNH	1DAB;Cla	n.d.	n.d.	L3 (1137-
NP_8865	Autotranspor	Bordetella	MPQRSQPF	GAVANSG	К	100.0	84	99.6	33	98.4	19	-	-	98.3	19	98.9	27	98.8	28	98.6	20	100.0	40	-	-	47.9	26	98.5	20	99.9	23	99.5	25	-	-	Autotransporter of unknc	IPR005546:	Yes	DCQTAGS/	1DAB;Cla	n.d.	n.d.	L1 (911-9
NP_8828	Autotranspor	Bordetella	MPQRSQPF	GAVANSG	К	100.0	84	99.6	33	98.5	16	-	-	98.2	21	98.9	23	98.8	30	98.8	20	100.0	40	-	-	60.6	26	99.0	21	100.0	24	99.6	27	-	-	Autotransporter of unknc	IPR005546:	Yes	DCQTVGS/	1DAB;Cla	n.d.	n.d.	L3 (447-5
YP_0032	Hypothetical	Cronobact	MLSSPMMV	NGDVTNC	Л	100.0	194	95.8	25	-	-	-	-	-	-	43.7	26	-	-	-	-	100.0	31	-	-	80.3	22	99.0	19	100.0	27	99.7	26	-	-	Autotransporter of unknc	IPR005546:	Yes	TNTITLNDS	1DAB;Cla	n.d.	265-383	L3 (229-3
YP_0014	Hypothetical	Cronobact	MDTSKTFR	NGDVTNC	Л	100.0	92	98.4	26	94.6	13	-	-	90.0	17	96.8	25	94.1	27	95.6	19	100.0	31	-	-	89.2	24	99.1	21	100.0	26	99.8	27	-	-	Autotransporter of unknc	IPR005546:	Yes	ETIISKDTS	1DAB;Cla	n.d.	428-554	L3 (252-3
YP_0033	Adhesin autc	Citrobacter	MVNKCFSLI	NGNVSNS	Л	100.0	101	97.6	23	23.9	11	-	-	41.7	15	93.6	24	69.6	23	92.9	15	100.0	29	-	-	76.4	15	99.0	14	99.9	26	99.7	25	-	-	Autotransporter of unknc	IPR005546:	Yes	ETIISESTS	n.d.	n.d.	265-383	L1 (57-15
YP_0036	Hypothetical	Enterobac	MQGNLINE	CNGNVTNC	Л	100.0	193	94.3	33	-	-	-	-	-	-	-	-	-	-	-	-	100.0	32	-	-	86.0	17	98.9	15	99.9	24	99.7	29	-	-	Autotransporter of unknc	IPR005546:	Yes	SDKNNTLV	n.d.	n.d.	200-322	T6 (234-2
YP_0033	Putative serir	Citrobacter	MNKIYSLKY	DSTVNNL	М	100.0	74	99.0	15	99.5	45	98.8	51	99.5	40	98.9	30	98.9	25	99.0	29	99.9	17	-	-	99.7	50	100.0	40	99.8	32	99.3	25	-	-	Protease / Peptidase S6	IPR005546:	Yes	QMDTRNF\	1DAB;Cla	n.d.	265-383	L1 (669-7
YP_6682	Protein sepa	Escherichi	MNKIYALKY	NSTAGNM	М	100.0	73	97.7	20	98.9	42	98.5	97	99.1	41	97.0	26	97.5	22	97.1	26	97.9	23	-	-	99.9	97	100.0	41	98.0	27	97.9	24	-	-	Protease / Peptidase S6	IPR005546:	Yes	TVGGNNPY	1DAB;Cla	n.d.	180-297	L1 (748-8
YP_0014	Hbp, Tsh	Escherichi	MNRIYSLRY	NSTAGNM	М	100.0	70	97.4	19	99.0	42	98.5	99	99.2	42	96.5	25	97.7	24	97.0	27	97.8	21	-	-	99.9	101	100.0	43	98.0	26	98.1	24	-	-	Protease / Peptidase S6	IPR005546:	Yes	TVNNELGY	1DAB;Cla	n.d.	428-554	L1 (749-8
YP_0015	Hypothetical	Salmonella	MNKIYALKY	DSRIGALF	М	100.0	75	98.7	11	99.4	46	98.9	54	99.5	39	98.6	27	98.6	20	98.7	22	97.7	13	-	-	99.9	56	100.0	40	97.7	32	97.6	22	-	-	Protease / Peptidase S6	IPR005546:	Yes	VVRYDIPY(	1DAB;Cla	n.d.	428-554	T3 (712-7
YP_0033	Putative serir	Citrobacter	MNKIYSLKY	DSGINNL	М	100.0	84	98.7	16	99.4	45	98.8	50	99.5	40	98.7	29	98.7	23	98.7	28	99.9	18	-	-	99.7	49	100.0	40	99.9	31	99.2	23	-	-	Protease / Peptidase S6	IPR005546:	Yes	IPYQLFRDF	1DAB;Cla	n.d.	265-383	T3 (677-7
NP_7522	Pic serine pro	Escherichi	MNKVYSLK	LWTLGGD	М	100.0	75	98.7	10	99.5	42	98.9	46	99.5	40	98.6	26	98.7	23	98.7	27	99.9	10	-	-	99.8	45	100.0	40	99.9	26	99.7	25	-	-	Protease / Peptidase S6	IPR005546:	Yes	IVRSDIAYQ	1DAB;Cla	n.d.	n.d.	L1 (806-8
NP_7087	Serine protea	Shigella fle	MNKVYSLK	DSAIHSLT	М	100.0	79	97.9	13	99.0	40	98.7	46	99.2	39	97.5	25	97.7	24	97.6	27	97.7	14	-	-	99.8	45	100.0	39	97.8	27	97.7	25	-	-	Protease / Peptidase S6	IPR005546:	Yes	IVRSDIAYQ	1DAB;Cla	n.d.	n.d.	T3 (701-7
AP_0032	Adhesin-like	Escherichi	MNRTSPYY	RVNSLLLI	Н	100.0	87	99.5	23	98.7	22	-	-	98.3	17	99.7	40	99.5	30	99.6	39	99.9	23	-	-	91.9	26	98.7	24	100.0	39	99.8	34	-	-	Self-Associating AutoTra	IPR030930:	Yes	AFTTNVIGV	1DAB;Cla	n.d.	n.d.	T4 (147-1
YP_0029	Adhesin-like	Escherichi	MNRTSPYY	RVNSLLI	Н	100.0	87	99.5	23	98.7	22	-	-	98.3	17	99.7	40	99.5	30	99.6	39	99.9	23	-	-	93.0	26	98.6	24	100.0	39	99.8	34	-	-	Self-Associating AutoTra	IPR030930:	Yes	AFTTNVIGV	1DAB;Cla	n.d.	473-603	T4 (147-1
YP_0017	Outer membr	Escherichi	MNRTSPYY	EYRVNSL	Н	100.0	81	99.5	22	98.8	23	-	-	98.4	22	99.7	38	99.5	34	99.6	38	99.9	23	-	-	95.1	29	98.9	21	100.0	36	99.9	32	-	-	Self-Associating AutoTra	IPR030930:	Yes	AFTTNVIGV	1DAB;Cla	n.d.	428-554	T4 (147-1
YP_0024	Adhesin-like	Escherichi	MNTIHLRCL	RVNSLLLI	Н	100.0	123	98.9	22	97.8	24	-	-	97.1	23	99.2	37	99.0	29	99.1	38	100.0	23	-	-	94.1	26	98.8	25	100.0	38	99.8	33	-	-	Self-Associating AutoTra	IPR030930:	Yes	AFTTNVIGV	1DAB;Cla	n.d.	473-603	T4 (137-1
YP_0017	Hypothetical	Escherichi	MNTIHLRCL	EYRVNSL	Н	100.0	125	99.2	22	98.5	23	-	-	98.2	21	99.5	39	99.3	30	99.4	38	100.0	22	-	-	95.3	27	99.0	21	100.0	36	99.9	28	-	-	Self-Associating AutoTra	IPR030930:	Yes	AFTPDVIGV	1DAB;Cla	n.d.	428-554	T4 (137-1
NP_3115	Abc transpor	Escherichi	MNTIHLRCL	RVNSLLLI	Н	100.0	127	99.5	22	98.7	23	-	-	98.2	22	99.7	37	99.6	34	99.6	38	99.9	23	-	-	94.3	27	98.8	21	100.0	35	99.8	32	-	-	Self-Associating AutoTra	IPR030930:	Yes	AFTPDVIGV	1DAB;Cla	n.d.	n.d.	T4 (137-1
YP_0034	Atp-binding c	Escherichi	MNRISRYYF	EYRVNSL	Н	100.0	68	54.8	19	-	-	-	-	-	-	94.6	38	91.3	29	92.1	35	99.9	22	-	-	94.2	27	98.8	23	100.0	35	99.8	33	-	-	Self-Associating AutoTra	IPR030930:	Yes	AFTSNVIGV	1DAB;Cla	n.d.	265-383	T4 (71-12
YP_4028	Atpase	Shigella dy	MAFRVLTM	CEYRVNS	Н	100.0	143	83.8	20	-	-	-	-	-	-	96.5	34	93.4	28	93.7	36	99.9	22	-	-	96.1	27	98.9	21	100.0	34	99.8	27	-	-	Self-Associating AutoTra	IPR030930:	Yes	TTRATRVT	1DAB;Cla	n.d.	180-297	T4 (155-2
YP_0015	Hypothetical	Salmonella	MVIFFSKTF	HSNGGW	Н	100.0	120	-	-	-	-	-	-	-	-	25.0	41	-	-	-	-	99.9	24	-	-	82.1	27	98.0	27	100.0	43	99.8	31	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADYNRNIID	1DAB;Cla	n.d.	428-554	T4 (196-2
YP_0039	Outer membr	Enterobac	MKSKIVGFI	YTLNSLTL	Н	100.0	59	29.5	18	-	-	-	-	-	-	95.3	35	82.3	20	90.7	37	99.9	25	-	-	94.2	30	98.5	23	100.0	37	99.8	33	-	-	Self-Associating AutoTra	IPR030930:	Yes	SDYNSRMI	1DAB;Cla	n.d.	91-203	T4 (108-1
YP_0017	Outer membr	Yersinia ps	MGGDNDG(	APFTFKRI	О	100.0	119	83.7	23	-	-	-	-	-	-	92.7	31	93.4	28	95.3	27	99.9	25	-	-	95.6	17	98.7	16	100.0	29	99.8	28	-	-	Autotransporter of unknc	IPR005546:	Yes	GAHTAQAI	1DAB;Cla	n.d.	428-554	n.d.
YP_0713	Pertactin fam	Yersinia ps	MKSRSNLN	LSAPFTFF	О	100.0	47	97.4	22	94.3	16	-	-	92.9	14	98.1	31	98.1	25	98.4	30	99.9	23	-	-	97.2	15	99.0	14	100.0	29	99.9	26	-	-	Autotransporter of unknc	IPR005546:	Yes	APSVAIPPF	1CII;Class	n.d.	91-203	T4 (424-4
YP_0018	Outer membr	Yersinia ps	MKSRSNLN	LSAPFTFF	О	100.0	50	97.3	23	92.8	15	-	-	94.0	13	97.8	30	97.9	26	98.2	28	99.9	23	-	-	96.9	14	99.0	14	100.0	28	99.9	25	-	-	Autotransporter of unknc	IPR005546:	Yes	APSVAIPPS	1CII;Class	n.d.	428-554	T4 (555-6
YP_0014	Autotranspor	Yersinia ps	MKSRSNLN	LSAPFTFF	О	100.0	50	97.2	21	94.4	17	-	-	94.3	14	97.9	30	97.9	26	98.1	29	99.9	22	-	-	96.0	16	98.7	15	100.0	29	99.8	25	-	-	Autotransporter of unknc	IPR005546:	Yes	APSVAIPPS	4CGK;Cla	n.d.	428-554	T4 (551-6
YP_0713	Pertactin fam	Yersinia ps	MKSRHHLN	LSAPFTFF	О	100.0	50	97.3	24	92.8	16	-	-	93.4	13	97.9	30	97.9	26	98.2	28	99.9	24	-	-	96.9	14	98.9	13	100.0	28	99.8	25	-	-	Autotransporter of unknc	IPR005546:	Yes	APSVAIPPF	1CII;Class	n.d.	91-203	T4 (605-6
YP_0713	Pertactin fam	Yersinia ps	MKNSNTLN	AHFALND	О	99.9	44	96.9	22	86.9	18	-	-	92.9	15	97.6	32	97.8	26	98.1	21	99.9	22	-	-	95.6	15	98.7	14	99.9	29	99.8	26	-	-	Autotransporter of unknc	IPR005546:	Yes	VPVVIDGGI	1CII;Class	n.d.	91-203	T4 (481-5
YP_0014	Autotranspor	Yersinia ps	MKNSNTLN	AQFALND	О	99.9	44	95.4	21	41.1	19	-	-	53.1	15	96.8	34	97.3	25	97.6	21	99.9	23	-	-	95.9	16	98.8	14	99.9	30	99.8	26	-	-	Self-Associating AutoTra	IPR030930:	Yes	VSQIATTD1	1CII;Class	n.d.	428-554	T4 (539-5
NP_9940	Autotranspor	Yersinia pe	MKNSNTLN	AHFALND	О	100.0	49	97.3	22	93.4	19	-	-	95.7	15	97.8	28	98.0	28	98.1	17	99.9	21	-	-	96.3	15	98.8	15	100.0	30	99.8	26	-	-	Self-Associating AutoTra	IPR030930:	Yes	VSQIATTD1	n.d.	n.d.	428-554	T4 (544-6
YP_6522	Putative auto	Yersinia pe	MKNSNTLN	AHFALND	О	99.9	48	97.1	15	73.3	18	-	-	82.5	15	97.2	34	97.6	27	97.8	16	99.9	21	-	-	96.0	16	98.8	14	99.9	30	99.8	26	-	-	Self-Associating AutoTra	IPR030930:	Yes	VSQIATTD1	n.d.	n.d.	180-297	T4 (544-6
NP_6686	Autotranspor	Yersinia pe	MKNSNTLN	AHFALND	О	99.9	48	98.1	14.0	73.3	18	-	-	82.5	15	97.2	34	97.6	27	97.8	16	99.9	23	-	-	95.7	17	98.8	15	99.9	31	99.8	28	-	-	Self-Associating AutoTra	IPR030930:	Yes	VSQIATTD1	n.d.	n.d.	n.d.	T4 (544-6
AP_0018	Predicted adl	Escherichi	MGIKQHNGI	NSNLDTL/	П	100.0	91	97.8	19	92.8	21	-	-	87.6	19	97.5	33	95.5	32	96.9	25	100.0	19	-	-	94.8	30	99.0	26	100.0	32	99.8	32	-	-	Autotransporter of unknc	IPR005546:	Yes	ADTGSQY(	1DAB;Cla	n.d.	n.d.	L1 (160-2
YP_0030	Outer membr	Escherichi	MGIKQHNGI	NSNLDTL/	П	100.0	93	98.9	20	98.2	21	-	-	97.5	20	98.8	33	98.4	35	98.7	27	100.0	22	-	-	91.2	25	98.7	24	100.0	30	99.7	35	-	-	Autotransporter of unknc	IPR005546:	Yes	ADTGSQY(	1DAB;Cla	n.d.	322-432	L1 (160-2
YP_0017	Adhesin	Escherichi	MGIKQHNGI	NSNLDTL/	П	100.0	91	97.8	19	92.8	21	-	-	87.6	19	97.5	33	95.5	32	96.9	25	100.0	19	-	-	91.5	29	98.8	23	100.0	31	99.8	34	-	-	Autotransporter of unknc	IPR005546:	Yes	ADTGSQY(	1DAB;Cla	n.d.	428-554	L1 (160-2
YP_0014	Outer membr	Escherichi	MGIKQHNGI	NSNLDTL/	П	100.0	97	98.9	20	98.0	21	-	-	97.0	19	98.6	37	98.2	34	98.5	24	100.0	23	-	-	97.4	30	99.3	28	100.0	32	99.8	32	-	-	Autotransporter of unknc	IPR005546:	Yes	ADTGSQY(	1DAB;Cla	n.d.	428-554	L1 (160-2
YP_0032	Adhesin	Escherichi	MGIKQHNGI	NSNLDTL/	П	100.0	97	98.6	21	97.3	21	-	-	96.0	19	98.3	33	97.7	34	98.1	25	100.0	20	-	-	97.9	29	99.5	28	100.0	33	99.8	35	-	-	Autotransporter of unknc	IPR005546:	Yes	ADTGSQY(	1DAB;Cla	n.d.	265-383	n.d.
YP_0032	Putative adhe	Escherichi	MGIKQHNGI	NSNLDTL/	П	100.0	97	98.6	21	97.3	21	-	-	96.0	19	98.3	33	97.7	34	98.1	25	99.9	18	-	-	95.8	31	98.9	26	100.0	32	99.8	34	-	-	Autotransporter of unknc	IPR005546:	Yes	ADTGSQY(	1DAB;Cla	n.d.	265-383	L1 (160-2
YP_0034	Outer membr	Escherichi	MGIKQHNGI	NSNLDTL/	П	100.0	97	98.5	21	97.5	21	-	-	96.8	19	98.3	33	97.9	34	98.1	25	100.0	18	-	-	97.7	32	99.4	29	100.0	32	99.8	33	-	-	Autotransporter of unknc	IPR005546:	Yes	ADTGSQY(	1DAB;Cla	n.d.	200-322	L1 (160-2
YP_6887	Adhesion and	Shigella fle	MGIKQHNGI	NSNLDTL/	П	100.0	91	98.8	17	98.1	21	-	-	97.4	19	98.6	33	98.3	32	98.6	25	100.0	19	-	-	94.7	30	98.8	24	100.0	32	99.8	31	-	-	Autotransporter of unknc	IPR005546:	Yes	ADTGSQY(	1DAB;Cla	n.d.	180-297	L1 (160-2
YP_0024	Putative adhe	Escherichi	MGIKQHNGI	NSNLDTL/	П	100.0	95	98.5	18	96.9	21	-	-	95.1	19	98.1	37	97.4	33	97.9	25	100.0	19	-	-	95.8	31	98.9	25	100.0	32	99.8	32	-	-	Autotransporter of unknc	IPR005546:	Yes	ADLGSQN(	1DAB;Cla	n.d.	473-603	L1 (160-2
YP_0033	Adhesin autc	Citrobacter	MEITAFRQF	SSNLDNL	П	100.0	202	96.2	16	-	-	-	-	-	-	92.7	31	61.5	23	84.1	32	100.0	20	-	-	95.1	27	98.9	28	100.0	30	99.8	33	-	-	Autotransporter of unknc	IPR005546:	Yes	EGISVGNG	1DAB;Cla	n.d.	265-383	L1 (150-2
YP_0033	Adhesin autc	Citrobacter	MKITAFRQF	SSNLDNL	П	100.0	75	97.3	18	93.4	21	-	-	85.1	22	97.4	33	95.5	34	96.6	32	99.9	19	-	-	90.9	26	98.6	28	100.0	31	99.7	34	-	-	Autotransporter of unknc	IPR005546:	Yes	ATIGSQTG\	1DAB;Cla	n.d.	265-383	L1 (151-2
YP_0033	Adhesin autc	Citrobacter	MKITAFRKF	SSNLDNL	П	100.0	79	98.5	18	97.4	22	-	-	96.4	22	98.4	32	98.0	34	98.1	32	99.9	20	-	-	92.6	26	98.7	28	100.0	31	99.7	36	-	-	Autotransporter of unknc	IPR005546:	Yes	ATIGSQTG\	1DAB;Cla	n.d.	265-383	n.d.
YP_0039	Outer membr	Pantoea va	MNNNTDLNI	TSNLDNL	П	100.0	89	97.5	17	94.6	19	-	-	86.0	14	97.6	31	96.4	29	96.9	34	100.0	18	-	-	92.4	27	98.8	17	100.0	29	99.8	35	-	-	Autotransporter of unknc	IPR005546:	Yes	AVVGSQR(	1DAB;Cla	n.d.	159-288	n.d.
YP_0015	Iga-specific s	Neisseria r	MVRDDVDY	WHLTGDS	Р	100.0	727	98.3	21	97.8	24	57.1	24	97.3	21	99.0	26	99.1	39	99.3	47	99.9	19	-	-	97.9	24	99.2	22	100.0	27	99.8	37	-	-	Protease / Peptidase S6	IPR005546:	Yes	GVETYKDK	1DAB;Cla	n.d.	428-554	T6 (749-8
NP_2737	Iga-specific s	Neisseria r	MKTKRFKIN	WHLTGDS	Р	100.0	609	98.3	21	98.0	25	77.5	18	97.6	20	99.1	26	99.2	39	99.4	47	99.9	20	-	-	97.1	23	98.9	21	100.0	26	99.8	36	-	-	Protease / Peptidase S6	IPR005546:	Yes	ALVRDDVD	1DAB;Cla	n.d.	n.d.	T6 (777-8
YP_9747	Iga1 protease	Neisseria r	MKTKRFKIN	WHLTGDS	Р	100.0	351	95.4	21	93.6	24	-	-	86.0	20	98.0	26	98.0	39	98.6	47	99.9	19	-	-	98.0	22	99.4	24	100.0	27	99.9	37	-	-	Protease / Peptidase S6	IPR005546:	Yes	ALVRDDVD	1DAB;Cla	n.d.	358-469	T6 (778-8
YP_2486	Iga-specific s	Haemophil	MLNKKFKLN	NSDVHQL	Р	100.0	531	99.0	24	98.8	22	95.5	28	98.6	19	99.4	28	99.5	31	99.6	61	99.9	20	-	-	98.0	26	99.4	21	100.0	28	99.9	32	-	-	Protease / Peptidase S6	IPR005546:	Yes	ALVRNDVD	1DAB;Cla	n.d.	180-297	T6 (854-9
YP_0012	Recombinatio	Haemophil	MLKNKKFKL	DSNVNQL	Р	100.0	540	98.9	23	98.8	22	95.8	20	98.5	19	99.3	28	99.4	32	99.6	74	99.9	21	-	-	98.4	26	99.5	21	100.0	29	99.9	31	-	-	Protease / Peptidase S6	IPR005546:	Yes	ALVRDDVD	1DAB;Cla	n.d.	428-554	T6 (807-8
YP_0041	Iga1 protease	Haemophil	MLNKKFKLN	DSNVNQL	Р	100.0	585	99.0	21	98.6	22	94.3	29	98.6	23	99.5	30	99.4	36	99.6	75	99.9	20	-	-	97.9	26	99.2	23	100.0	30	99.9	36	-	-	Protease / Peptidase S6	IPR005546:	Yes	ALVRDDVD	1DAB;Cla	n.d.	91-203	T6 (807-8
NP_4391	Immunoglobi	Haemophil	MLNKKFKLN	DSNVNQL	Р	100.0	450	96.7	20	96.5	21	-	-	95.5	21	98.3	30	98.3	33	98.8	96	99.9	19	-	-	98.1	31	99.4	27	100.0	29	99.9	35	-	-	Protease / Peptidase S6	IPR005546:	Yes	ALVRDDVD	1DAB;Cla	n.d.	n.d.	T6 (804-8
YP_0041	Iga1 protease	Haemophil	MLNKKFKLN	LTGDSNV	Р	100.0	530	99.1	24	99.0	26	96.8	26	98.8	21	99.5	27	99.5	38	99.7	73	100.0	24	-	-	96.2	21	98.9	18	100.0	27	99.9	36	-	-	Protease / Peptidase S6	IPR005546:	Yes	ALVRDDVD	1DAB;Cla	n.d.	91-203	T6 (797-8
YP_0029	Pertactin fam	Bartonella	MQRKLKLSF	GNTETYIK	С	100.0	167	97.8	19	96.3	21	-	-	35.6	20	98.9	29	98.5	24	98.2	21	99.9	21	-	-	95.5	19	98.6	17	100.0	28	99.8	24	-	-	Autotransporter of unknc	IPR005546:	Yes	DGVATPKE	1DAB;Cla	n.d.	473-603	T4 (144-1
YP_0016	Hypothetical	Bartonella	MQRKLKLSF	DKSTHIQ(	С	100.0	169	98.5	18	98.1	25	-	-	90.9	21	99.4	32	99.1	30	99.1	29	99.9	17	-	-	93.9	14	98.4	18	100.0	31	99.8	29	-	-	Self-Associating AutoTra	IPR030930:	Yes	DGVATLKE	1DAB;Cla	n.d.	428-554	T4 (144-1
YP_0333	Hypothetical	Bartonella	MDKNIINGS	DHQTTVE	С	100.0	265	76.7	6	-	-	-	-	-	-	-	-	-	-	44.1	10	100.0	24	-	-	96.9	10	98.3	11	100.0	18	99.7	13	99.8	12.0	Self-Associating AutoTra	IPR030930:	Yes	KGETLGQC	1DAB;Cla	n.d.	91-203	T4 (108-1
YP_0016	Hypothetical	Bartonella	MLSSFFVQ/	SSLIQKLS	С	100.0	55	97.9	22	96.7	19	-	-	91.4	16	98.6	31	97.8	23	97.7	26	99.9	23	-	-	58.6	21	98.4	23	100.0	31	99.8	21	-	-	Autotransporter of unknc	IPR005546:	Yes	SLFCQSSS	1DAB;Cla	n.d.	428-554	T4 (160-2
YP_0029	Pertactin fam	Bartonella	MALQYKWK	SSLIQKLS	С	100.0	61	99.2	21	99.0	20	25.4	16	98.1	19	99.6	30	99.3	21	99.3	28	99.9	17	-	-	92.0	15	98.6	22	100.0	31	99.8	25	-	-	Autotransporter of unknc	IPR005546:	Yes	AAANEGGI	1DAB;Cla	n.d.	473-603	T4 (174-2
YP_0333	Hypothetical	Bartonella	MRYKWKRS	SSLIQKLS	С	100.0	59	98.6	18	98.0	17	-	-	96.1	20	99.2	31	98.7	24	98.7	24	99.9	24	-	-	56.9	20	97.9	21	100.0	32	99.7	22	-	-	Autotransporter of unknc	IPR005546:	Yes	GEEGRSWI	1DAB;Cla	n.d.	91-203	T4 (176-2
YP_0041	Hypothetical	Bartonella	MQYKWKLG	TDSASIKN	С	100.0	59	97.6	18	86.6	17	-	-	50.2	18	97.7	32	96.1	29	96.6	18	99.9	21	-	-	85.7	17	98.3	20	100.0	29	99.7	29	-	-	Self-Associating AutoTra	IPR030930:	Yes	NKNEIQSF(	1DAB;Cla	n.d.	91-203	T4 (157-2
YP_9888	Putative adhe	Bartonella	MQYKWKLS	FTQIENLS	С	100.0	57	98.9	17	97.9	16	-	-	94.0	16	99.1	27	98.5	27	98.7	21	99.9	19	-	-	79.4	18	98.0	19	100.0	27	99.7	24	-	-	Autotransporter of unknc	IPR005546:	Yes	SDWCDVVI	1DAB;Cla	n.d.	352-483	T4 (140-1
YP_0041	Hypothetical	Bartonella	MQCKLVSG	TDSASIKN	С	99.9	57	97.1	19	86.5	17	-	-	35.7	15	96.7	33	89.3	25	89.1	19	99.9	21	-	-	78.3	18	98.1	20	99.9	29	99.7	30	-	-	Autotransporter of unknc	IPR005546:	Yes	DVERFNVA	1DAB;Cla	n.d.	91-203	T4 (73-13
YP_0029	Pertactin fam	Bartonella	MQYKHKLSI	IEIEDLGG	С	99.9	180	94.2	5	-	-	-	-	-	-	74.6	10	60.2	9	85.6	11	99.9	17	-	-	96.1	14	98.1	20	99.9	36	99.8	27	99.9	25.0	Autotransporter of unknc	IPR005546:	Yes	VENRNGEL	1DAB;Cla	n.d.	473-603	T4 (72-13
YP_9888	Putative adhe	Bartonella	MKNEYRLVI	GASSIHIT	С	100.0	102	90.6	16	58.3	16	-	-	-	-	97.4	27	94.2	25	93.1	24	99.9	20	-	-	94.0	19	98.1	17	100.0	26	99.8	25	-	-	Self-Associating AutoTra	IPR030930:	Yes	EEKKQGGF	1DAB;Cla	n.d.	352-483	T4 (58-11
YP_0011	Autotranspor	Yersinia pe	MNANADSC	AYATVDN	С	100.0	164	98.4	21	97.9	15	-	-	96.7	13	98.8	40	98.4	24	98.5	21	99.9	23	-	-	94.1	22	98.8	14	100.0	39	99.8	30	-	-	Autotransporter of unknc	IPR005546:	Yes	MVDQIFVT	1DAB;Cla	n.d.	428-554	n.d.
YP_0014	Autotranspor	Yersinia ps	MNTNSKKT	AYATVDN	С	100.0	58	98.2	21	97.2	13	-	-	95.5	13	98.4	39	98.1	31	98.1	22	100.0	22	-	-	92.5	23	98.7	14	100.0	38	99.8	31	-	-	Autotransporter of unknc	IPR005546:	Yes	ANADSCTD	1DAB;Cla	n.d.	428-554	n.d.
NP_2749	Serine-type p	Neisseria r	MRFTHTTPF	SSHTGAL	Т	100.0	211	99.2	24	98.2	19	-	-	97.6	23	99.4	31	99.5	45	99.3	31	100.0	23	-	-	98.4	20	99.4	19	100.0	37	99.9	46	-	-	Protease / Peptidase S6	IPR005546:	Yes	SIVRNDVD	1DAB;Cla	n.d.	n.d.	n.d.
YP_2091	Adhesion and	Neisseria g	MKTTDKRT	GNLNLDN	Т	100.0	116	98.7	23	97.4	17	-	-	96.7	23	99.1	38	99.4	74	98.9	37	99.9	18	-	-	95.2	19	98.4	19	100.0	32	99.9	54	-	-	Protease / Peptidase S6	IPR005546:	Yes	GHTYFGIN	1DAB;Cla	n.d.	91-203	T3 (798-8
YP_9758	Iga-specific s	Neisseria r	MKTTDKRT	GNLNLDN	Т	100.0	108	98.9	21	97.2	19	32.4	20	95.6	18	99.1	39	99.4	75	98.9	36	100.0	20	-	-	97.3	22	98.9	20	100.0	32	99.9	58	-	-	Protease / Peptidase S6	IPR005546:	Yes	GHTYFGIN	1DAB;Cla	n.d.	352-483	T3 (798-8
YP_0015	Adhesion and	Neisseria r	MKTTDKRT	GNLNLDN	Т	100.0	160	99.0	23	97.8	19	30.6	21	96.8	19	99.2	41	99.5	63	99.1	36	99.9	20	-	-	96.4	21	98.8	20	100.0	31	99.9	58	-	-	Protease / Peptidase S6	IPR005546:	Yes	GHTYFGIN	1DAB;Cla	n.d.	428-554	T3 (798-8
YP_0030	Iga-specific s	Neisseria r	MKTTDKRT	GNLNLDN	Т	100.0	164	98.9	22	97.3	24	54.9	21	96.1	18	99.1	39	99.4	71	98.9	32	99.9	22	-	-	97.6	18	99.0	20	100.0	32	99.9	54	-	-	Protease / Peptidase S6	IPR005546:	Yes	GHTYFGIN	1DAB;Cla	n.d.	265-383	L1 (826-9
YP_0040	Iga-specific s	Neisseria l	MKTTDKRT	GNLNLDN	Т	99.9	394	95.1	25	-	-	-	-	22.0	19	96.4	35	98.1	72	95.6	31	99.9	27	-	-	90.4	18	98.1	20	99.9	37	99.8	72	-	-	Protease / Peptidase S6	IPR005546:	Yes	GHTYFGIN	1DAB;Cla	n.d.	91-203	L1 (792-9
YP_0023	Iga-specific s	Neisseria r	MKTTDKRT	GNLNLDN	Т	100.0	165	98.9	22	97.1	23	29.9	20	96.5	18	99.1	37	99.4	60	99.1	34	100.0	19	-	-	95.8	23	98.5	20	100.0	32	100.0	57	-	-	Protease / Peptidase S6	IPR005546:	Yes	GHTYFGIN	1DAB;Cla	n.d.	473-603	T3 (791-8
YP_0020	Hypothetical	Neisseria g	MILTSNINQ(	GNLNLDN	Т	100.0	197	98.8	22	97.3	19	24.2	17	96.2	19	99.0	38	99.4	71	99.0	31	100.0	18	-	-	96.4	20	98.5	17	100.0	30	100.0	58	-	-	Protease / Peptidase S6	IPR005546:	Yes	LNADNQFN	1DAB;Cla	n.d.	132-263	T3 (394-4
YP_2479	Adhesion and	Haemophil	MKKTVFRLN	DTTLQNL	Т	100.0	164	99.2	24	98.2	25	44.7	25	97.4	22	99.5	39	99.6	97	99.4	29	100.0	21	-	-	95.3	25	98.4	18	100.0	34	100.0	79	-	-	Protease / Peptidase S6	IPR005546:	Yes	GHTYFGID	1DAB;Cla	n.d.	180-297	L1 (758-8
YP_0033	Outer membr	Streptobac	MVKGKYMT	VIVSDKFL	У	100.0	33	99.9	22	99.6	12	95.2	16	99.5	17	99.8	19	99.8	22	99.7	26	100.0	23	-	-	93.5	18	99.1	17	100.0	21	99.8	21	-	-	Autotransporter of unknc	IPR005546:	Yes	KTVDEKINE	1DAB;Cla	IPR011050;IPF	265-383	n.d.
YP_0033	Outer membr	Streptobac	MVKNKYMG	EDKFVNK	У	100.0	36	99.9	29	99.8	18	97.6	22	99.6	18	99.8	27	99.8	25	99.7	20	100.0	30	-	-	95.6	20	99.0	17	99.9	26	99.7	25	-	-	Autotransporter of unknc	IPR005546:	Yes	ETVDEKINE	1DAB;Cla	n.d.	265-383	n.d.

YP_0033	Outer membr	Streptobac	MCINRSVFLASNIEVEL	Y	100.0	35	99.8	28	99.6	17	94.3	21	99.5	17	99.7	24	99.7	25	99.6	21	100.0	28	-	-	96.1	19	99.1	17	100.0	24	99.7	25	-	-	Autotransporter of unknc	IPR005546:	Yes	ESIDEKINE	1DAB;Cla	n.d.	265-383	n.d.
YP_0033	Outer membr	Streptobac	MNINKTIFLFEDKFVNK	Y	100.0	35	99.7	31	99.4	18	85.4	20	99.3	17	99.5	26	99.5	26	99.3	21	100.0	32	-	-	95.3	20	99.0	18	99.9	27	99.7	26	-	-	Autotransporter of unknc	IPR005546:	Yes	KTIDEKINE	1DAB;Cla	n.d.	265-383	n.d.
YP_0033	Outer membr	Streptobac	MYRKGIFLF EASNIEVE	Y	100.0	36	99.9	28	99.7	17	96.8	18	99.6	17	99.8	24	99.8	25	99.7	21	100.0	28	-	-	95.7	19	98.9	17	99.9	24	99.7	25	-	-	Autotransporter of unknc	IPR005546:	Yes	EQSTEKIVE	1DAB;Cla	n.d.	265-383	n.d.
YP_0014	Outer membr	Xanthobac	MFFRLVQG GNSAVGL	-	93.1	95	88.1	20	-	-	-	-	-	-	-	-	-	-	-	-	89.6	31	-	-	-	-	-	-	93.1	30	-	-	-	-	Autotransporter of unknc	IPR005546:	Yes	ACNNAINLI	2AGM;Cla	n.d.	428-554	L1 (921-9
YP_0021	Hypothetical	Burkholder	MGSNKKRA GGLVVSG	-	96.0	145	49.7	13	-	-	-	-	-	-	-	-	21.4	16	-	-	96.0	18	-	-	-	-	81.0	15	91.7	18	65.5	16	-	-	Autotransporter of unknc	IPR005546:	Yes	ACVGTTTV	n.d.	IPR011050;IPF	132-263	L3 (334-4
YP_8403	Outer membr	Burkholder	MRMRMGSN QLGQPLK	-	95.6	143	42.6	13	-	-	-	-	-	-	-	-	-	-	-	-	95.6	18	-	-	-	-	67.1	18	90.5	14	19.1	15	-	-	Autotransporter of unknc	IPR005546:	Yes	ACTGTTIVS	n.d.	IPR011050;IPF	345-479	L3 (378-4
YP_0017	Outer membr	Burkholder	MRMRMGSN TVTGTGT	-	96.3	118	36.4	17	-	-	-	-	-	-	-	-	33.4	16	-	-	96.1	20	-	-	-	-	88.8	17	96.3	19	86.6	15	-	-	Autotransporter of unknc	IPR005546:	Yes	ACTGTTTV	n.d.	IPR011050;IPF	428-554	L3 (378-4
YP_0015	Outer membr	Burkholder	MRMRMGSN LVVSGGG	-	97.3	118	69.5	16	-	-	-	-	-	-	-	-	-	-	-	-	97.3	18	-	-	-	-	81.3	17	96.2	14	55.9	16	-	-	Autotransporter of unknc	IPR005546:	Yes	ACTGAPTV	n.d.	IPR011050;IPF	428-554	L3 (89-17
YP_0020	Outer membr	Stenotroph	MNRIYRLVF TITVGGNL	-	95.4	86	69.9	15	-	-	-	-	-	-	-	-	-	-	-	-	95.4	19	-	-	14.4	21	90.3	19	80.5	14	18.1	17	-	-	Autotransporter of unknc	IPR005546:	Yes	ACNTVGST	n.d.	IPR011050;IPF	132-263	L1 (419-5
YP_0015	Hypothetical	Azorhizobi	MARRKSDLI GGLTVNA	-	97.6	118	33.4	22	-	-	-	-	-	-	-	-	-	-	-	-	97.6	22	-	-	-	-	75.0	8	96.3	15	31.8	15	-	-	Autotransporter of unknc	IPR005546:	Yes	AREVRADS	n.d.	n.d.	428-554	L1 (659-7
YP_0036	Outer membr	Burkholder	MSLTGLGSI GGSLVNT	-	96.7	117	63.1	19	-	-	-	-	25.0	9	-	-	-	-	-	-	96.7	27	-	-	1.7	18	75.9	15	86.1	17	58.5	13	-	-	Autotransporter of unknc	IPR005546:	Yes	DGAGGAA	n.d.	n.d.	200-322	L3 (1150-
YP_0039	Outer membr	Achromob	MQPLSPIDV GVINAGLI	-	100.0	37	58.8	22	-	-	-	-	-	-	-	-	-	-	-	-	100.0	31	-	-	25.5	30	98.6	18	99.9	36	99.6	30	-	-	Autotransporter of unknc	IPR005546:	Yes	QTIDGGSV	n.d.	n.d.	91-203	L3 (112-1
NP_8892	Autotranspor	Bordetella	MKPIFRIEH GTVGDTT	-	96.7	108	89.6	17	-	-	-	-	-	-	64.6	16	20.5	18	-	-	96.7	15	-	-	-	-	52.7	18	95.9	11	36.2	18	-	-	Autotransporter of unknc	IPR005546:	Yes	ELLGVDGG	1DAB;Cla	IPR011050	446-577	n.d.
YP_4121	Outer membr	Nitrosospir	MAKRKKSA GDSGASG	-	97.1	109	94.8	17.0	-	-	-	-	31.2	15	-	-	-	-	-	-	97.1	18	-	-	2.7	12	59.7	14	94.8	19	12.2	20	-	-	Autotransporter of unknc	IPR005546:	Yes	QQGLMGD	1DAB;Cla	IPR011050	180-297	n.d.
YP_0035	VirG protein	Yersinia ps	MFMPFREFI SGNLISAC	-	100.0	74	99.9	42	99.3	16	-	-	99.1	17	99.7	30	99.6	33	99.6	26	100.0	44	-	-	8.8	23	98.5	22	99.9	26	99.5	29	-	-	Autotransporter of unknc	IPR005546:	Yes	SCNNSGG	1DAB;Cla	IPR011050;IPF	200-322	n.d.
YP_6491	Outer membr	Yersinia ps	MFMPFREFI SGNLISAC	-	100.0	74	99.9	42	99.3	16	-	-	99.1	17	99.7	30	99.6	33	99.6	26	100.0	44	-	-	8.8	23	98.5	22	99.9	26	99.5	29	-	-	Autotransporter of unknc	IPR005546:	Yes	SCNNSGG	1DAB;Cla	IPR011050;IPF	180-297	n.d.
YP_0717	Autotranspor	Yersinia ps	MFMPFREFI SGNLISAC	-	100.0	76	100.0	41	99.7	17	83.1	25	99.6	19	99.8	31	99.8	33	99.8	27	100.0	44	-	-	23.3	23	98.7	19	99.9	24	99.6	28	-	-	Autotransporter of unknc	IPR005546:	Yes	SCNNSGG	1DAB;Cla	IPR011050;IPF	91-203	n.d.
YP_0018	Outer membr	Yersinia ps	MFMPFREFI SGNLISAC	-	100.0	76	99.9	41	99.4	18	30.5	27	99.2	14	99.7	31	99.6	30	99.6	27	100.0	43	-	-	11.2	25	98.6	20	99.9	22	99.6	28	-	-	Autotransporter of unknc	IPR005546:	Yes	SCNNSGG	1DAB;Cla	IPR011050;IPF	428-554	n.d.
YP_0013	Autotranspor	Yersinia ps	MPFRAFISS SGNLISAC	-	100.0	74	99.9	43	99.3	18	31.7	25	99.1	19	99.6	30	99.6	30	99.5	27	100.0	43	-	-	33.5	25	98.8	19	99.9	27	99.7	26	-	-	Autotransporter of unknc	IPR005546:	Yes	TCTSSGVG	1DAB;Cla	IPR011050;IPF	428-554	n.d.
NP_9918	Hypothetical	Yersinia ps	MHIQFREPI SGNLISAC	-	100.0	81	99.5	44	98.4	14	-	-	98.3	15	98.8	19	98.7	27	98.6	22	100.0	42	-	-	15.5	22	98.6	17	99.9	25	99.4	27	-	-	Autotransporter of unknc	IPR005546:	Yes	ACNSSGV(	1DAB;Cla	IPR011050;IPF	428-554	n.d.
YP_0689	Autotranspor	Yersinia ps	MHIQFREPI SGNLISAC	-	100.0	81	99.4	44	98.2	14	-	-	98.1	14	98.5	28	98.5	32	98.4	22	100.0	44	-	-	17.5	24	98.4	18	99.9	27	99.4	32	-	-	Autotransporter of unknc	IPR005546:	Yes	ACNSSGV(	1DAB;Cla	IPR011050;IPF	91-203	n.d.
YP_0014	Autotranspor	Yersinia ps	MHIQFREPI SGNLISAC	-	100.0	81	99.9	41	99.6	16	21.0	24	99.4	18	99.8	31	99.8	29	99.7	27	100.0	43	-	-	48.4	20	98.8	18	99.9	25	99.6	26	-	-	Autotransporter of unknc	IPR005546:	Yes	ACNSSGV(	1DAB;Cla	IPR011050;IPF	428-554	n.d.
YP_0018	Outer membr	Yersinia ps	MHIQFREPI SGNLISAC	-	100.0	81	99.6	44	98.7	16	-	-	98.6	15	99.0	21	99.0	25	98.8	22	100.0	42	-	-	10.9	23	98.7	17	99.9	26	99.5	30	-	-	Autotransporter of unknc	IPR005546:	Yes	ACNSSGV(	1DAB;Cla	IPR011050;IPF	428-554	n.d.
YP_0017	Outer membr	Yersinia ps	MHIQFREPI SGNLISAC	-	100.0	84	99.5	44	98.3	14	-	-	98.0	15	98.7	20	98.7	31	98.6	23	100.0	43	-	-	45.7	24	98.7	17	99.9	27	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	ACNSSGV(	1DAB;Cla	IPR011050;IPF	428-554	n.d.
YP_6492	Hypothetical	Yersinia ps	MHIQFREPI SGNLISAC	-	100.0	81	99.5	44	98.4	14	-	-	98.3	15	98.8	19	98.7	27	98.6	22	100.0	43	-	-	26.3	25	98.4	16	99.9	27	99.5	32	-	-	Autotransporter of unknc	IPR005546:	Yes	ACNSSGV(	1DAB;Cla	IPR011050;IPF	180-297	n.d.
YP_6538	Hypothetical	Yersinia ps	MHIQFREPI SGNLISAC	-	100.0	81	99.5	44	98.4	14	-	-	98.3	15	98.8	19	98.7	27	98.6	22	100.0	43	-	-	26.3	25	98.4	16	99.9	27	99.5	32	-	-	Autotransporter of unknc	IPR005546:	Yes	ACNSSGV(	1DAB;Cla	IPR011050;IPF	180-297	n.d.
YP_0023	Hypothetical	Yersinia ps	MNPITLAIIV SGNLISAC	-	100.0	50	99.8	42	99.2	15	-	-	99.0	14	99.6	29	99.5	28	99.4	23	100.0	44	-	-	79.4	22	98.9	19	99.9	29	99.7	31	-	-	Autotransporter of unknc	IPR005546:	Yes	ACNSSGV(	1DAB;Cla	IPR011050;IPF	473-603	n.d.
YP_0035	Hypothetical	Yersinia ps	MHIQFREPI SGNLISAC	-	100.0	81	99.5	44	98.4	14	-	-	98.3	15	98.8	19	98.7	27	98.6	22	100.0	43	-	-	26.3	25	98.4	16	99.9	27	99.5	32	-	-	Autotransporter of unknc	IPR005546:	Yes	ACNSSGV(	1DAB;Cla	IPR011050;IPF	200-322	n.d.
NP_6675	Hypothetical	Yersinia ps	MHIQFREPI SGNLISAC	-	100.0	84	99.1	44	97.8	14	-	-	97.5	15	98.2	24	98.1	32	98.0	21	100.0	43	-	-	16.1	25	98.7	15	99.9	26	99.5	32	-	-	Autotransporter of unknc	IPR005546:	Yes	ACNSSGV(	1DAB;Cla	IPR011050;IPF	n.d.	n.d.
NP_9931	Hypothetical	Yersinia ps	MLKHFRLGI SGNLISAC	-	100.0	81	99.9	43	99.6	19	53.4	25	99.5	19	99.8	31	99.8	32	99.8	27	100.0	43	-	-	26.3	20	98.7	19	99.9	25	99.6	27	-	-	Autotransporter of unknc	IPR005546:	Yes	ACTSPGV(	1DAB;Cla	IPR011050;IPF	428-554	n.d.
YP_6517	Hypothetical	Yersinia ps	MLKHFRLGI SGNLISAC	-	100.0	81	99.9	43	99.6	19	53.4	25	99.5	19	99.8	31	99.8	32	99.8	27	100.0	42	-	-	43.9	20	98.7	16	99.9	27	99.6	26	-	-	Autotransporter of unknc	IPR005546:	Yes	ACTSPGV(	1DAB;Cla	IPR011050;IPF	180-297	n.d.
YP_0035	Outer membr	Yersinia ps	MLKHFRLGI SGNLISAC	-	100.0	81	99.9	43	99.6	19	53.4	25	99.5	19	99.8	31	99.8	32	99.8	27	100.0	42	-	-	43.9	20	98.7	16	99.9	27	99.6	26	-	-	Autotransporter of unknc	IPR005546:	Yes	ACTSPGV(	1DAB;Cla	IPR011050;IPF	200-322	n.d.
YP_0011	Hypothetical	Yersinia ps	MLKHFRLGI SGNLISAC	-	100.0	81	99.9	43	99.6	19	53.4	25	99.5	19	99.8	31	99.8	32	99.8	27	100.0	42	-	-	43.9	20	98.7	16	99.9	27	99.6	26	-	-	Autotransporter of unknc	IPR005546:	Yes	ACTSPGV(	1DAB;Cla	IPR011050;IPF	428-554	n.d.
YP_0023	Hypothetical	Yersinia ps	MLKHFRLGI SGNLISAC	-	100.0	81	99.9	43	99.3	14	20.3	25	99.0	11	99.6	28	99.6	30	99.5	22	100.0	43	-	-	26.3	20	98.7	19	99.9	25	99.6	27	-	-	Autotransporter of unknc	IPR005546:	Yes	ACTSPGV(	1DAB;Cla	IPR011050;IPF	473-603	n.d.
YP_6478	Hypothetical	Yersinia ps	MLKHFRLGI SGNLISAC	-	100.0	81	99.9	43	99.6	19	53.4	25	99.5	19	99.8	31	99.8	32	99.8	27	100.0	42	-	-	43.9	20	98.7	16	99.9	27	99.6	26	-	-	Autotransporter of unknc	IPR005546:	Yes	ACTSPGV(	1DAB;Cla	IPR011050;IPF	180-297	n.d.
YP_0016	Outer membr	Yersinia ps	MLKHFRLGI SGNLISAC	-	100.0	81	99.9	43	99.6	19	54.4	24	99.5	19	99.8	31	99.8	33	99.8	27	100.0	43	-	-	26.3	20	98.7	19	99.9	25	99.6	27	-	-	Autotransporter of unknc	IPR005546:	Yes	ACTSPGV(	1DAB;Cla	IPR011050;IPF	428-554	n.d.
YP_0018	Outer membr	Yersinia ps	MYIQFREPI SGNLISAC	-	100.0	81	99.9	44	99.5	17	-	-	99.3	19	99.7	31	99.7	30	99.6	26	100.0	44	-	-	79.4	22	98.9	19	99.9	29	99.7	31	-	-	Autotransporter of unknc	IPR005546:	Yes	SCNSSGV(	1DAB;Cla	IPR011050;IPF	132-263	n.d.
YP_0717	Autotranspor	Yersinia ps	MYIQFREPI SGNLISAC	-	100.0	81	99.9	43	99.4	19	-	-	99.1	13	99.7	30	99.7	29	99.6	26	100.0	44	-	-	79.4	22	98.9	19	99.9	29	99.7	31	-	-	Autotransporter of unknc	IPR005546:	Yes	SCNSSGV(	1DAB;Cla	IPR011050;IPF	91-203	n.d.
YP_0017	Outer membr	Yersinia ps	MYIQFREPI SGNLISAC	-	100.0	81	99.9	42	99.3	16	-	-	99.1	14	99.7	31	99.7	32	99.6	27	100.0	43	-	-	76.2	22	98.9	20	100.0	28	99.8	25	-	-	Autotransporter of unknc	IPR005546:	Yes	SCNSSGG(	1DAB;Cla	IPR011050;IPF	428-554	n.d.
YP_0013	Autotranspor	Yersinia ps	MYIQFREPI SGNLISAC	-	100.0	81	99.9	43	99.4	14	-	-	99.2	12	99.7	31	99.7	29	99.6	25	100.0	43	-	-	48.4	20	98.8	18	99.9	25	99.6	26	-	-	Autotransporter of unknc	IPR005546:	Yes	SCNSSGV(	1DAB;Cla	IPR011050;IPF	428-554	n.d.
NP_6691	Hypothetical	Yersinia ps	MILLMILMLI SGNLISAC	-	100.0	43	99.7	42	99.0	14	-	-	98.7	11	99.3	28	99.3	29	99.2	25	100.0	42	-	-	44.0	23	98.8	20	99.9	25	99.7	26	-	-	Autotransporter of unknc	IPR005546:	Yes	NNSGSGG	1DAB;Cla	IPR011050;IPF	n.d.	n.d.
NP_3962	Autotranspor	Agrobacter	MTVPAAFFE YSFSGDV	-	100.0	134	99.5	37	97.7	10	-	-	96.8	7	98.8	24	98.6	25	98.4	18	100.0	39	-	-	76.0	16	98.9	9	100.0	21	99.8	27	-	-	Autotransporter of unknc	IPR005546:	Yes	GSCDPAGF	1DAB;Cla	IPR011050;IPF	n.d.	n.d.
YP_0022	Outer membr	Klebsiella	MKLNALANS TTTLSNV(	-	100.0	56	98.1	44	36.0	15	-	-	60.2	12	94.1	22	88.1	24	93.8	17	100.0	46	-	-	87.6	19	98.9	16	100.0	31	99.7	26	-	-	Autotransporter of unknc	IPR005546:	Yes	ADPCKSTD	1DAB;Cla	IPR011050;IPF	132-263	T5 (261-3
YP_0035	Hypothetical	Pantoea an	MSRTTRKF VTLAGKVI	-	100.0	40	97.8	33	48.1	17	-	-	80.5	18	94.7	22	88.2	25	92.9	17	100.0	35	-	-	85.7	20	99.1	18	100.0	24	99.8	29	-	-	Autotransporter of unknc	IPR005546:	Yes	VCSQEGNH	1DAB;Cla	IPR011050;IPF	200-322	T3 (469-5
NP_5385	Serine protea	Brucella m	MVDGYVSG SLDNAGL	-	100.0	107	99.2	35	96.7	19	-	-	97.0	19	98.5	30	98.0	26	98.1	19	100.0	35	-	-	91.8	19	99.0	20	100.0	30	99.8	28	-	-	Autotransporter of unknc	IPR005546:	Yes	AISTPKMKS	1DAB;Cla	n.d.	n.d.	L3 (740-8
NP_6980	Outer membr	Brucella su	MLLHSRWK SLDNAGL	-	100.0	55	99.8	35	98.9	19	-	-	98.7	19	99.5	30	99.3	26	99.2	17	100.0	36	-	-	88.0	17	98.9	23	100.0	31	99.8	28	-	-	Autotransporter of unknc	IPR005546:	Yes	AQTKYGSC	1DAB;Cla	n.d.	n.d.	L3 (957-1
YP_0031	Outer membr	Brucella m	MLLHSRWK SLDNAGL	-	100.0	55	99.7	36	98.3	21	-	-	98.2	19	99.3	30	99.1	30	99.0	19	100.0	36	-	-	81.4	16	99.0	20	100.0	31	99.8	28	-	-	Autotransporter of unknc	IPR005546:	Yes	AQTKYGSC	1DAB;Cla	n.d.	265-383	n.d.
YP_0013	Outer membr	Ochrobact	MLKRLNGKI ALSNGGL	-	100.0	67	99.8	35	98.7	23	-	-	98.7	19	99.4	30	99.3	36	99.2	24	100.0	35	-	-	82.3	18	99.2	20	100.0	30	99.8	35	-	-	Autotransporter of unknc	IPR005546:	Yes	MTPVLAQT	1DAB;Cla	n.d.	428-554	L3 (560-6
YP_0013	Outer membr	Ochrobact	MSGRGGAS AALSNGG	-	100.0	125	100.0	35	99.7	22	-	-	99.6	20	99.9	30	99.8	36	99.8	24	100.0	35	-	-	80.2	20	99.2	20	100.0	31	99.8	36	-	-	Autotransporter of unknc	IPR005546:	Yes	VDIVGNNN	1DAB;Cla	n.d.	428-554	L3 (233-3
YP_0013	Outer membr	Ochrobact	MHSRLNSSI ALSNGGL	-	100.0	86	100.0	35	99.6	23	-	-	99.5	20	99.9	30	99.8	34	99.8	24	100.0	35	-	-	82.3	18	99.2	20	100.0	30	99.8	35	-	-	Autotransporter of unknc	IPR005546:	Yes	TDYTAPVT	1DAB;Cla	n.d.	428-554	L3 (266-3
YP_0034	Outer membr	Xenorhabc	MKICKQALN ALSNNGR	-	100.0	79	99.8	41	98.8	16	-	-	98.4	15	99.4	25	99.2	30	99.2	22	100.0	42	-	-	81.0	23	99.0	16	100.0	26	99.7	29	-	-	Autotransporter of unknc	IPR005546:	Yes	ASSCGSSN	1DAB;Cla	n.d.	265-383	n.d.
NP_6993	Outer membr	Brucella su	MGAWSGAC ALYNAGT	-	100.0	65	99.3	35	96.8	16	-	-	97.0	13	98.4	22	98.1	30	98.0	19	100.0	35	-	-	85.5	20	99.0	16	100.0	24	99.8	35	-	-	Autotransporter of unknc	IPR005546:	Yes	ANFTNNST	1DAB;Cla	n.d.	n.d.	L3 (306-4
YP_0031	Outer membr	Brucella m	MDGFASCG ALYNAGT	-	100.0	89	99.8	34	98.8	19	-	-	98.7	15	99.4	25	99.3	30	99.2	21	100.0	34	-	-	88.7	19	99.1	20	100.0	25	99.8	32	-	-	Autotransporter of unknc	IPR005546:	Yes	GGQGGYG	1DAB;Cla	n.d.	265-383	L3 (520-6
YP_0013	Outer membr	Ochrobact	MSKLYTLFL SLGNTGN	-	100.0	74	98.8	34	94.6	14	-	-	94.0	16	97.6	32	96.7	30	96.5	21	100.0	37	-	-	71.5	19	98.8	20	100.0	31	99.7	30	-	-	Autotransporter of unknc	IPR005546:	Yes	DDLPGSAC	1DAB;Cla	n.d.	428-554	L3 (488-5
YP_0013	Autotranspor	Janthinob	MRVVIKINL SLSNAGT	-	100.0	77	99.8	37	99.0	11	-	-	99.0	17	99.4	21	99.4	30	99.2	19	100.0	37	-	-	66.2	16	99.1	16	100.0	21	99.8	29	-	-	Autotransporter of unknc	IPR005546:	Yes	QNSIFTGNI	1DAB;Cla	n.d.	428-554	T4 (73-13
YP_0013	Autotranspor	Janthinob	MTINIARVR SLSNAGT	-	100.0	72	99.4	35	97.9	12	-	-	97.8	18	98.7	21	98.5	30	98.3	17	100.0	36	-	-	86.1	17	99.0	19	100.0	20	99.7	31	-	-	Autotransporter of unknc	IPR005546:	Yes	QNYTGNDI	1DAB;Cla	n.d.	428-554	T4 (90-14
YP_0019	Outer membr	Brucella at	MPNLANQD GATTNSG	-	100.0	213	100.0	40	99.8	12	-	-	99.7	13	99.9	27	99.9	30	99.9	21	100.0	40	-	-	53.4	18	98.8	14	100.0	25	99.7	31	-	-	Autotransporter of unknc	IPR005546:	Yes	GTNGLTKA	1DAB;Cla	n.d.	132-263	L3 (495-5
YP_2238	Outer membr	Brucella at	MPNLANQD GATTNSG	-	100.0	213	100.0	40	99.8	12	-	-	99.7	13	99.9	27	99.9	30	99.9	21	100.0	41	-	-	82.0	18	98.8	17	99.9	26	99.7	32	-	-	Autotransporter of unknc	IPR005546:	Yes	GTNGLTKA	1DAB;Cla	n.d.	91-203	L3 (495-5
YP_4192	Aminoacyl-tr	Brucella m	MPNLANQD GATTNSG	-	100.0	213	100.0	40	99.8	12	-	-	99.7	13	99.9	27	99.9	30	99.9	21	100.0	41	-	-	82.0	18	98.8	17	99.9	26	99.7	32	-	-	Autotransporter of unknc	IPR005546:	Yes	GTNGLTKA	1DAB;Cla	n.d.	180-297	TS (192-2
YP_0031	Outer membr	Brucella m	MPNLANQD GATTNSG	-	100.0	213	100.0	39	99.7	13	-	-	99.7	14	99.9	27	99.9	29	99.8	22	100.0	40	-	-	73.9	17	98.9	14	100.0	26	99.8	31	-	-	Autotransporter of unknc	IPR005546:	Yes	GTNGLTKA	1DAB;Cla	n.d.	265-383	L3 (495-5
YP_0012	Outer membr	Brucella ov	MPNLANQD GATTNSG	-	100.0	213	100.0	38	99.7	11	-	-	99.7	14	99.9	26	99.9	28	99.8	22	100.0	39	-	-	60.0	17	98.8	14	100.0	25	99.7	31	-	-	Autotransporter of unknc	IPR005546:	Yes	GTNGLTKA	1DAB;Cla	n.d.	428-554	L3 (495-5
NP_7003	Outer membr	Brucella su	MPNLANQD GATTNSG	-	100.0	214	100.0	39	99.7	13	-	-	99.7	14	99.9	27	99.9	30	99.9	22	100.0	40	-	-	54.8	18	98.8	14	100.0	26	99.7	32	-	-	Autotransporter of unknc	IPR005546:	Yes	GTNGLTKA	1DAB;Cla	n.d.	n.d.	L3 (495-5
NP_5411	Extracellular	Brucella m	MLQTGTGTI GATTNSG	-	100.0	199	100.0	36	99.6	11	-	-	99.5	13	99.8	27	99.8	30	99.8	21	100.0	41	-	-	96.0	17	99.2	13	100.0	25	99.8	30	-	-	Autotransporter of unknc	IPR005546:	Yes	DGGQVSSL	1DAB;Cla	n.d.	n.d.	L3 (206-3
YP_6731	Outer membr	Mesorhizol	MMGTGTAA GTVCDTN	-	100.0	177	100.0	39	99.7	16	-	-	99.6	14	99.9	29	99.8	25	99.8	23	100.0	36	-	-	88.8	18	98.9	14	100.0	26	99.7	26	-	-	Autotransporter of unknc	IPR005546:	Yes	GENSSIGA	1DAB;Cla	n.d.	180-297	L1 (2017-
YP_6760	Outer membr	Mesorhizol	MKNPKNAIS WLDNAGI	-	100.0	115	99.8	37	99.0	19	-	-	98.9	18	99.5	31	99.4	28	99.3	22	100.0	39	-	-	88.8	23	98.9	18	100.0	30	99.7	30	-	-	Autotransporter of unknc	IPR005546:	Yes	GSYFVSNE	1DAB;Cla	n.d.	180-297	L3 (583-6
YP_6759	Outer membr	Mesorhizol	MQKPICVKF KTARLQG	-	100.0	85	99.3	35	97.4	14	-	-	96.8	13	98.4	22	98.1	26	98.1	21	100.0	38	-	-	97.1	20	99.2	14	100.0	22	99.8	26	-	-	Autotransporter of unknc	IPR005546:	Yes	ADFTASTQ	1DAB;Cla	n.d.	180-297	L3 (382-4
YP_7668	Adhesin prec	Rhizobium	MSPFCGSP DVTNAGV	-	100.0	90	99.2	37	96.8	17	-	-	97.2	19	98.2	27	97.9	31	97.9	20	100.0	43	-	-	86.3	21	98.8	18	99.9	25	99.6	25	-	-	Autotransporter of unknc	IPR005546:	Yes	ADWRGLTS	1DAB;Cla	n.d.	180-297	L3 (126-2
YP_0039	Outer membr	Achromob	MVARYCLR PGPRSGF	-	100.0	87	95.8	37	-	-	-	-	-	-	39.0	31	-	-	-	-	100.0	37	-	-	89.0	25	98.9	17	99.9	33	99.7	32	-	-	Autotransporter of unknc	IPR005546:	Yes	QSVSSSGC	1DAB;Cla	n.d.	91-203	L3 (86-16
NP_8881	Autotranspor	Bordetella	MVGRSCHR ANKGTIAF	-	100.0	65	67.2	31	-	-	-	-	-	-	-	-	-	-	-	-	100.0	33	-	-	88.0	25	98.9	17	100.0	27	99.7	30	-	-	Autotransporter of unknc	IPR005546:	Yes	QSVSGFGC	1DAB;Cla	n.d.	262-400	L3 (61-14
NP_8808	Autotranspor	Bordetella	MVGRSCHR ANKGTIAF	-	100.0	65	65.3	37	-	-	-	-	-	-	-	-	-	-	-	-	100.0	35	-	-	75.7	22	98.9	17	100.0	28	99.7	31	-	-	Autotransporter of unknc	IPR005546:	Yes	QSVSGSGC	1DAB;Cla	n.d.	n.d.	L3 (220-3
YP_0013	Autotranspor	Janthinob	MTISNTDFL TNSGVIAF	-	100.0	105	98.8	41	95.9	17	-	-	95.5	18	97.5	28	97.1	26	97.0	21	100.0	41	-	-	84.5	22	99.0	16	100.0	22	99.7	27	-	-	Autotransporter of unknc	IPR005546:	Yes	QSIIGNGDL	1DAB;Cla	n.d.	428-554	L1 (350-4
YP_0013	Outer membr	Ochrobact	MKHSTGTD SFSNAGT	-	100.0	91	98.1	42	84.2	12	-	-	84.0	13	96.3	27	92.8	31	94.9	26	100.0	38	-	-	74.8	17	98.7	17	100.0	25	99.8	28	-	-	Autotransporter of unknc	IPR005546:	Yes	DDYHYIGTI	1DAB;Cla	n.d.	428-554	L3 (327-4
YP_0013	Autotranspor	Janthinob	MHCSNKCF LSLDNGG	-	100.0	109	98.5	34	85.7	10	-	-	84.7	14	96.7	21	96.1	25	95.1	18	100.0	35	-	-	86.9	16	99.0	16	100.0	21	99.8	29	-	-	Autotransporter of unknc	IPR005546:	Yes	QVFIENDE	1DAB;Cla	n.d.	428-554	L1 (81-15
YP_0034	Outer membr	Xenorhabc	MLRIYRAVL ALNNSGR	-	100.0	69	99.7	38	98.5	17	-	-	98.2	21	99.2	24	99.0	25	98.9	22	100.0	38	-	-	88.8	23	99.1	22	100.0	26	99.7	25	-	-	Autotransporter of unknc	IPR005546:	Yes	ASTCGPTIS	n.d.	n.d.	265-383	n.d.
YP_0019	Outer membr	Brucella at	MTVEQGTA SLDNAGL	-	100.0	176	90.7	35	-	-	-	-	-	-	46.1	27	-	-	-	-	100.0	36	-	-	48.5	18	98.6	20	100.0	31	99.7	27	-	-	Autotransporter of unknc	IPR005546:	Yes	TVDLGGFD	1DAB;Cla	n.d.	132-263	n.d.
YP_6491	Outer membr	Yersinia ps	MGGLVHGG GGILLSNT	-	100.0	181	99.4	40	98.3	15	-	-	98.0	11	98.7	21	98.7	24	98.4	20	100.0	42	-	-	-	-	98.7	16	99.9	26	99.5	31	-	-	Autotransporter of unknc	IPR005546:	Yes	ADIAADPTL	1DAB;Cla	IPR011050;IPF	180-297	n.d.
NP_6707	Adhesin	Yersinia ps	MGGLVHGG GGILLSNT	-	100.0	181	99.5	39	98.4	11	-	-	98.2	11	98.9	20	98.8	23	98.5	20	100.0	42	-	-	5.8	23	98.6	13	99.9	26	99.5	31	-	-	Autotransporter of unknc	IPR005546:	Yes	ADIAADPTL	1DAB;Cla	IPR011050;IPF	n.d.	n.d.
YP_6532	Outer membr	Yersinia ps	MGGLVHGG GGILLSNT	-	100.0	181	99.3	41	98.1	15	-	-	97.9	11	98.6	20	98.5	24	98.2	20	100.0	43	-	-	8.1	23	98.6	13	99.9	25	99.4	30	-	-	Autotransporter of unknc	IPR005546:	Yes	ADIAADPTL	1DAB;Cla	IPR011050;IPF	180-297	n.d.
YP_0035	Hypothetical	Yersinia ps	MHGGNGTA GGILLSNT	-	100.0	177	99.3	44	98.3	11	-	-	97.9	11	98.5	24	98.6	25	98.2	22	100.0	42	-	-	7.9	23	98.8	17	99.9	25	99.5	28	-	-	Autotransporter of unknc	IPR005546:	Yes	ADIAADPTL	1DAB;Cla	IPR011050;IPF	200-322	n.d.
YP_0017	Adhesin	Escherichi	MIASLFSAN TVNGNLT	-	100.0	61	99.6	29	98.1	22	-	-	97.9	17	99.1	25	99.2	27	99.0	24	100.0	33	-	-	93.9	15	98.8	18	99.9	26	99.8	27	-	-	Autotransporter of unknc	IPR005546:	Yes	CQGYDVK	1DAB;Cla	n.d.	428-554	L3 (350-4
AP_0028	Adhesin	Escherichi	MRIIFLRKE TVNGNLT	-	100.0	72	99.6	31	98.5	21	-	-	98.4	17	99.3	25	99.3	27	99.2	29	100.0	32	-	-	92.5	19	99.0	21	100.0	28	99.8	28	-	-	Autotransporter of unknc	IPR005546:	Yes	AVTDSCQC	1DAB;Cla	n.d.	n.d.	n.d.
YP_0029	Adhesin	Escherichi	MRIIFLRKE TVNGNLT	-	100.0	72	99.6	31	98.5	21	-	-	98.4	17	99.3	25	99.3	27	99.2	29	100.0	32	-	-	69.7	15	98.4	20	99.9	28	99.7	29	-	-	Autotransporter of unknc	IPR005546:	Yes	AVTDSCQC	1DAB;Cla	n.d.	473-603	L3 (366-4
YP_8533	Adhesin	Escherichi	MNMRIISLR SDVTVNG	-	100.0	75	99.7	31	98.7	23	-	-	98.6	17	99.5	23	99.5	29	99.3	25	100.0	32	-	-	92.8	16	98.8	19	99.9	24	99.8	31	-	-	Autotransporter of unknc	IPR005546:	Yes	AVIDSCQG	1DAB;Cla	n.d.	345-479	n.d.
YP_0019	Yaph protein	Xanthomo	MNRAFALV LVSSG	-	97.2	185	23.7	16	-	-	-	-	-	-	-	-	-	-	-	-	97.2	20	-	-	-	-	45.7	9	93.8	13	3.3	8	-	-	Autotransporter of unknc	IPR005546:	Yes	NASSTAVS	n.d.	IPR011050;IPF	91-203	L3 (181-2
YP_3638	Filamentous	Xanthomo	MNRAFALV LVSSGTV	-	97.0	202	51.0	18	-	-	-	-	-	-	-	-	-	-	-	-	97.0	20	-	-	-	-	46.6	13	95.3	19	6.3	7	-	-	Autotransporter of unknc	IPR005546:	Yes	NASSTAVT	n.d.	IPR011050;IPF	180-297	L1 (849-9
YP_2432	Yaph protein	Xanthomo	MRARRSLL PVTVGPG	-	97.9	120	33.7	15	-	-	-	-	-	-	-	-	-	-	-	-	97.9	16	-	-	8.5	16	72.2	14	96.8	13	60.9	12	-	-	Autotransporter of unknc	IPR005546:	Yes	WGACTPAI	n.d.	IPR011050;IPF	180-297	L1 (422-5
NP_6373	Yaph protein	Xanthomo	MRARRSLL PVTVGPG	-	97.8	119	33.7	15	-	-	-	-	-	-	-	-	-	-	-	-	97.8	15	-	-	2.4	17	61.6	15	95.8	14	28.5	10	-	-	Autotransporter of unknc	IPR005546:	Yes	WGACTPAI	n.d.	IPR011050;IPF	n.d.	T1 (416-4
YP_0019	Filamentous	Xanthomo	MRARRSLL PVTVGPG	-	97.5	119	39.2	15	-	-	-	-	-	-	-	-	-	-	-	-	97.5	16	-	-	2.6	17	41.8	14	95.8	15	16.5	13	-	-	Autotransporter of unknc	IPR005546:	Yes	WGACTPAI	n.d.	IPR011050;IPF	132-263	T1 (416-4
YP_3494	Outer membr	Pseudomo	MEARFGVAI GSVVNHG	-	96.3	118	69.6	18	-	-	-	-	-	-	-	-	-	-	-	-	96.3	18	-	-	5.4	16	87.1	19	90.6	18	48.0	13	-	-	Autotransporter of unknc	IPR005546:	Yes	TCSAAGST	n.d.	n.d.	180-297	L1 (795-8
YP_6066	Outer membr	Pseudomo	MDRRAASR GGLVNNG	-	94.7	164	46.2	19	-	-	-	-	20.0	11	-	-	-	-	-	-	94.7	22	-	-	-	-	39.0	18	45.0	20	13.7	20	-	-	Autotransporter of unknc	IPR005546:	Yes	GNNNTLNN	n.d.	n.d.	180-297	L1 (1236-
YP_0021	Shda	Salmonella	MNKTYNIW TFTVNGD	-	100.0	91	99.9	29	99.5	17	-	-	99.5	18	99.8	24	99.7	28	99.7	25	100.0	29	-	-	82.2	15	99.0	20	100.0	23	99.8	27	-	-	Autotransporter of unknc	IPR005546:	Yes	EETRVSIPS	1DAB;Cla	n.d.	132-263	L1 (448-5
NP_4614	C-terminal re	Salmonella	MRNMYVVA FTLNGDL	-	100.0	137	99.7	40	98.6	14	-	-	98.4	19	99.1	28	99.1	31	98.9	21	100.0	41	-	-	93.3	20	98.7	20	99.9	29	99.7	31	-	-	Autotransporter of unknc	IPR005546:	Yes	DNQVSYAE	2ZJ6;Clas	n.d.	n.d.	L1 (162-2
YP_0015	Hypothetical	Salmonella	MRNMYVVA FTLNGDLI	-	100.0	137	99.7	38	98.5	13	-	-	98.4	18	99.1	26	99.1	31	98.9	21	100.0	37	-	-	95.3	20	98.8	18	99.9	26	99.8	31	-	-	Autotransporter of unknc	IPR005546:	Yes	DNQVSYAE	1DAB;Cla	n.d.	428-554	L1 (162-2
YP_0021	Shda	Salmonella	MNRTYSIVV FTLNGDLI	-	100.0	148	99.6	41	98.6	14	-	-	98.5	18	99.1	27	99.1	31	98.9	24	100.0	39	-	-	93.7	20	98.7	22	99.9	27	99.7	31	-	-	Autotransporter of unknc	IPR005546:	Yes	DNPVSYID(	1DAB;Cla	n.d.	132-263	L1 (150-2
YP_0022	Host colonisz	Salmonella	MRNMYVVA FTLNGDLI	-	100.0	118	99.7	38	98.7	13	-	-	98.5	17	99.1	24	99.1	30	98.9	21	100.0	38	-	-	88.2	20	98.5	20	99.9	27	99.7	31	-	-	Autotransporter of unknc	IPR005546:	Yes	DNQVSYAE	1DAB;Cla	n.d.	132-263	L3 (162-2
YP_0026	Aida autotrar	Salmonella	MYVVASEL FTLNGDL	-	100.0	114	99.5	38	98.1	13	-	-	98.0	17	98.9	24	98.8	25	98.6	22	100.0	38	-	-	89.8	19	98.5	17	99.9	25	99.7	29	-	-	Autotransporter of unknc	IPR005546:	Yes	DNQVSYAE	1DAB;Cla	n.d.	473-603	L1 (92-17
YP_0018	Outer membr	Yersinia ps	MNTIFKVIWI TGNVENH	-	100.0	95	100.0	37	99.8	11	-	-	99.6	13	99.9	27	99.8	25	99.8	18	100.0	36	-	-	86.4	19	98.9	15	100.0	26	99.7	27	-	-	Autotransporter of unknc	IPR005546:	Yes	ADITVSTQ	1DAB;Cla	IPR011050;IPF	132-263	L3 (342-4
YP_0717	Pertactin fam	Yersinia ps	MNTIFKVIWI TGNVENH	-	100.0	95	100.0	36	99.8	11	-	-	99.7	14	99.9	28	99.9	27	99.9	18	100.0	36	-	-	86.3	20	99.0	16	100.0	27	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	ADITVSTQ	1DAB;Cla	IPR011050;IPF	91-203	L3 (342-4
YP_6532	Putative auto	Yersinia ps	MNTIFKVIWI TGNVENY	-	100.0	95	100.0	36	99.8	11	62.4	22	99.7	14	99.9	28	99.8	27	99.9	18	100.0	36	-	-	86.0	19	99.0	15	100.0	27	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	ADITVSTQ	1DAB;Cla	IPR011050;IPF	180-297	L3 (342-4
YP_0023	Autotranspor	Yersinia ps	MNTIFKVIWI TGNVENY	-	100.0	95	100.0	36	99.8	11	40.5	19	99.7	14	99.9	28	99.9	27	99.9	18	100.0	36	-	-	83.3	20	98.9	14	100.0	27	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	ADITVSTQ	1DAB;Cla	IPR011050;IPF	473-603	L2 (393-4
NP_9946	Autotranspor	Yersinia ps	MLRLYMNTI TGNVENY	-	100.0	100	100.0	36	99.8	11	-	-	99.7	14	99.9	28	99.9	28	99.9	18	100.0	36	-	-	86.0	19	99.0	15	100.0	27	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	ADITVSTQ	1DAB;Cla	IPR011050;IPF	428-554	L2 (398-4
NP_6706	Autotranspor	Yersinia ps	MLRLYMNTI TGNVENY	-	100.0	100	100.0	36	99.8	11	-	-	99.7	14	99.9	28	99.8	27	99.8	18	100.0	36	-	-	86.0	19	99.0	15	100.0	27	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	ADITVSTQ	1DAB;Cla	IPR011050;IPF	n.d.	L2 (398-4
YP_6491	Autotranspor	Yersinia ps	MNTIFKVIWI TGNVENY	-	100.0	95	100.0	36	99.8	11	40.5	19	99.7	14	99.9	28	99.9	27	99.9	18	100.0	36	-	-	84.9	20	98.7	14	100.0	27	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	ADITVSTQ	1DAB;Cla	IPR011050;IPF	180-297	L3 (342-4
YP_0035	Putative auto	Yersinia ps	MNTIFKVIWI TGNVENY	-	100.0	95	100.0	36	99.8	11	40.5	19	99.7	14	99.9	28	99.9	27	99.9	18	100.0	36	-	-	84.9	20	98.8	16	100.0	27	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	ADITVSTQ	1DAB;Cla	IPR011050;IPF	200-322	L2 (393-4
YP_0013	Autotranspor	Yersinia ps	MLRLCMNAI TGNVENH	-	100.0	96	100.0	37	99.8	11	-	-	99.7	13	99.9	28	99.9	27	99.9	18	100.0	36	-	-	86.3	20	99.0	16	100.0	27	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	IDAQGCTD	1DAB;Cla	n.d.	428-554	L1 (482-5
YP_0017	Outer membr	Yersinia ps	MNTIFKVIWI TGNVENH	-	100.0	89	100.0	36	99.8	11	65.6	19	99.6	13	99.9	27	99.8	27	99.8	18	100.0	36	-	-	86.4	19	98.9	15	100.0	26	99.7	27	-	-	Autotransporter of unknc	IPR005546:	Yes	ADITVSTQ	1DAB;Cla	IPR011050;IPF	428-554	L3 (342-4
YP_0017	Outer membr	Escherichi	MNKIYKVIW SLTNTGVI	-	100.0	74	99.9	36	99.2	21	41.9	25	99.1	20	99.7	26	99.6	33	99.6	29	100.0	36	-	-	96.5	22	99.1	21	100.0	30	99.8	34	-	-	Autotransporter of unknc	IPR005546:	Yes	DVILSPDW	1DAB;Cla	n.d.	428-554	L1 (760-8
YP_0014	Outer membr	Citrobacte	MNKVFKVIW PVSFTVG	-	100.0	94	99.9	43	99.3	14	-	-	99.1	18	99.6	25	99.6	28	99.6	18	100.0	43	-	-	91.7	16	98.9	20	100.0	28	99.8	28	-	-	Autotransporter of unknc	IPR005546:	Yes	ADLQVYDF	1DAB;Cla	IPR011050;IPF	428-554	L1 (891-9
NP_887C	Autotranspor	Bordetella	MIAGVSSH GDVANRC	-	100.0	147	99.9	40	99.6	17	34.8	22	99.5	17	99.8	23	99.7	25	99.7	21	100.0	40	-	-	65.6	23	98.7	17	99.9	21	99.5	25	-	-	Autotransporter of unknc	IPR005546:	Yes	AWANCTTS	1DAB;Cla	n.d.	n.d.	L3 (687-7
YP_0032	Aida-i adhesi	Escherichi	MNKIYRLKV ASVNYTL(	-	100.0	104	99.8	43	98.8	20	-	-	98.4	19	99.3	28	99.1	29	99.1	19	100.0	43	-	-	88.9	21	98.9	21	100.0	33	99.7	29	-	-	Autotransporter of unknc	IPR005546:	Yes	DDVIVNQD	1DAB;Cla	IPR011050;IPF	265-383	L3 (119-2
YP_0024	Aida-i adhesi	Escherichi	MAFNALLFI ASVNYTL(	-	100.0	127	99.8	43	98.6	21	-	-	98.2	20	99.2	28	98.9	29	99.0	19	100.0	43	-	-	88.9	21	98.9	21	100.0	33	99.7	29	-	-	Autotransporter of unknc	IPR005546:	Yes	DDVIVNQD	1DAB;Cla	IPR011050;IPF	473-603	L3 (141-2
YP_0023	Aida-i adhesi	Escherichi	MAFNALLFI ASVNYTL(	-	100.0	127	99.8	43	98.8	21	-	-	98.7	20	99.3	28	99.2	29	99.1	19	100.0	43	-	-	88.9	21	98.9	21	100.0	33	99.7	29	-	-	Autotransporter of unknc	IPR005546:	Yes	DDVIVNQD	1DAB;Cla	IPR011050;IPF	473-603	TS (115-1
YP_0022	Outer membr	Escherichi	MAFNALLFI ASVNYAL(	-	100.0	127	99.8	43	98.7	22	-	-	98.3	20	99.2	28	99.0	28	99.0	19	100.0	43	-	-	85.9	20	98.7	21	99.9	29	99.7	27	-	-	Autotransporter of unknc	IPR005546:	Yes	DDVIVNQD	1DAB;Cla	IPR011050;IPF	476-592	L3 (141-2
NP_286C	Beta-barrel o	Escherichi	MAFNALLFI ASVNYAL(	-	100.0	126	99.8	43	98.7	22	-	-	98.3	20	99.2	28	99.0	28	99.0	19	100.0	43	-	-	85.8	20	98.7	18	99.9	30	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	DDVIVNQD	1DAB;Cla	IPR011050;IPF	n.d.	L3 (141-2
YP_0030	Beta-barrel o	Escherichi	MAFNALLFI ASVNYAL(	-	100.0	127	99.8	43	98.7	22	-	-	98.3	20	99.2	28	99.0	28	99.0	19	100.0	43	-	-	85.8	20	98.7	18	99.9	30	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	DDVIVNQD	1DAB;Cla	IPR011050;IPF	265-383	TS (115-1
YP_0034	Aida-i adhesi	Escherichi	MAFNALLFI ASVNYAL(	-	100.0	126	99.8	43	98.7	22	-	-	98.3	20	99.2	28	99.0	28	99.0	19	100.0	43	-	-	85.8	20	98.7	18	99.9	30	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	DDVIVNQD	1DAB;Cla	IPR011050;IPF	265-383	TS (115-1
NP_3083	Aida-i adhesi	Escherichi	MNKIYRLKV ASVNYAL(	-	100.0	106	99.6	43	98.2	21	-	-	97.9	20	99.0	28	98.6	28	98.7	19	100.0	43	-	-	85.9	20	98.7	21	99.9	29	99.7	27	-	-	Autotransporter of unknc	IPR005546:	Yes	DDVIVNQD	1DAB;Cla	IPR011050;IPF	n.d.	L3 (119-2
YP_0032	Aida-i adhesi	Escherichi	MNKIYRLKV ASVNYTL(	-	100.0	106	99.8	44	98.9	20	-	-	98.4	19	99.3	28	99.1	28	99.1	20	100.0	44	-	-	88.7	21	98.7	20	99.9	30	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	DDVIVNQD	1DAB;Cla	IPR011050;IPF	265-383	L3 (119-2
YP_0032	Aida-i adhesi	Escherichi	MNKIYRLKV ASVNYTL(	-	100.0	69	99.7	44	98.5	21	-	-	98.3	20	99.0	28	98.8	30	98.8	19	100.0	44	-	-	88.7	21	98.7	21	99.9	30	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	DDVIVNQD	1DAB;Cla	IPR011050;IPF	265-383	TS (93-11
YP_0014	Outer membr	Escherichi	MNKIYRLKV ASVNYTL(	-	100.0	104	99.8	44	98.6	21	-	-	98.3	20	99.2	28	99.0	29	99.0	19	100.0	44	-	-	88.7	21	98.7	20	99.9	30	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	DDVIVNQD	1DAB;Cla	IPR011050;IPF	428-554	TS (93-11
YP_0014	Outer membr	Escherichi	MNKIYRLKV SVHYTLG	-	100.0	105	99.7	44	98.5	21	-	-	98.3	20	99.1	28	98.8	31	98.8	20	100.0	44	-	-	94.3	22	98.9	21	100.0	31	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	DDVIVNQD	1DAB;Cla	IPR011050;IPF	428-554	L3 (119-2
YP_0023	Aida-i adhesi	Escherichi	MDVLLEIVM PVNYTLG	-	100.0	63	99.8	44	98.7	20	-	-	98.5	19	99.3	29	99.0	30	99.0	19	100.0	44	-	-	91.9	22	98.7	19	99.9	33	99.7	29	-	-	Autotransporter of unknc	IPR005546:	Yes	DDVIVKKDI	1DAB;Cla	IPR011050;IPF	473-603	L3 (127-2

YP_0024	Putative lipoç	Escherichi	MNRIYRVIW	TVGGNLE	-	100.0	103	99.7	35	98.4	14	-	-	98.2	22	99.1	29	99.1	34	98.8	22	100.0	35	-	-	88.9	21	98.9	24	99.9	27	99.8	34	-	-	Autotransporter of unknc	IPR005546:	Yes	ASLEVDNG	1DAB;Cla	n.d.	473-603	L3 (250-3
YP_0017	Putative lipoç	Escherichi	MNRIYRVIW	TTIGGNLE	-	100.0	103	99.6	35	98.4	14	-	-	98.0	22	99.1	29	99.0	34	98.8	21	100.0	34	-	-	89.0	21	98.9	25	99.9	26	99.8	34	-	-	Autotransporter of unknc	IPR005546:	Yes	ASLEVDNG	1DAB;Cla	n.d.	428-554	L3 (146-2
YP_0024	Lipoprotein/a	Escherichi	MNRIYRVIW	TTIGGNLE	-	100.0	104	99.6	35	98.1	14	-	-	97.6	22	99.0	29	98.8	34	98.6	22	100.0	34	-	-	89.0	21	98.9	25	99.9	26	99.8	34	-	-	Autotransporter of unknc	IPR005546:	Yes	ASLEVDNI	1DAB;Cla	n.d.	473-603	n.d.
YP_0032	Autotranspor	Escherichi	MNRIYRVIW	TTIGGNLE	-	100.0	104	99.5	35	97.8	14	-	-	97.4	21	98.8	27	98.7	34	98.4	21	100.0	35	-	-	91.0	21	99.0	25	100.0	26	99.8	32	-	-	Autotransporter of unknc	IPR005546:	Yes	ASLEVDND	1DAB;Cla	n.d.	265-383	n.d.
YP_0022	Putative lipoç	Escherichi	MNRIYRVIW	TTIGGNLE	-	100.0	104	99.5	35	97.8	14	-	-	97.4	21	98.8	27	98.7	34	98.4	21	100.0	34	-	-	79.2	22	98.9	25	100.0	30	99.8	36	-	-	Autotransporter of unknc	IPR005546:	Yes	ASLEVDND	1DAB;Cla	n.d.	473-603	n.d.
YP_0023	Putative lipoç	Escherichi	MNRIYRVIW	TTIGGNLE	-	100.0	100	99.7	35	98.5	14	-	-	98.1	21	99.2	29	99.1	34	98.9	21	100.0	35	-	-	83.9	18	98.9	25	100.0	26	99.8	34	-	-	Autotransporter of unknc	IPR005546:	Yes	ASLEVDND	1DAB;Cla	n.d.	473-603	n.d.
YP_0039	Outer membr	Enterobact	MNKIFRVVR	NLSSAGT	-	100.0	89	99.7	32	98.7	18	-	-	98.6	20	99.3	30	99.3	31	99.0	26	100.0	32	-	-	90.8	20	99.1	24	100.0	27	99.8	34	-	-	Autotransporter of unknc	IPR005546:	Yes	ATVAIDAG	1DAB;Cla	n.d.	91-203	n.d.
YP_0014	Outer membr	Escherichi	MNRIYRVIW	FDMENGC	-	100.0	103	99.7	34	98.7	14	-	-	98.7	21	99.2	27	99.2	30	98.9	21	100.0	34	-	-	94.6	22	99.2	23	100.0	25	99.8	32	-	-	Autotransporter of unknc	IPR005546:	Yes	ASLEVDND	1DAB;Cla	n.d.	428-554	L3 (250-3
NP_8583	IcsA	Shigella fle	MNQIHKFFC	DPVIVTGN	-	100.0	53	99.6	100	97.9	14	-	-	97.7	12	98.7	22	98.3	26	98.3	18	100.0	100	-	-	85.5	20	98.9	15	99.9	25	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	TPLSGTQE	1DAB;Cla	n.d.	n.d.	n.d.
YP_3133	Icsa/virg	Shigella sc	MNQIHKFFC	DPVIVTGN	-	100.0	105	99.8	100	99.0	17	-	-	98.8	12	99.4	22	99.2	26	99.2	17	100.0	100	-	-	84.3	19	98.9	15	100.0	25	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	IPLSGTQEL	1DAB;Cla	n.d.	180-297	n.d.
YP_4062	Icsa (plasmic	Shigella dy	MNQIHKFFC	DPVIVTGN	-	100.0	72	99.8	100	98.9	18	-	-	98.7	12	99.3	22	99.1	26	99.1	20	100.0	100	-	-	83.8	20	98.9	14	100.0	25	99.7	29	-	-	Autotransporter of unknc	IPR005546:	Yes	IPLSGTQEL	1DAB;Cla	n.d.	180-297	n.d.
YP_0018	Outer membr	Shigella bo	MNQIHKFFC	DPVIVTGN	-	100.0	108	99.8	99	98.8	17	-	-	98.4	12	99.2	22	99.0	26	98.9	19	100.0	99	-	-	81.2	19	98.8	15	99.9	25	99.7	29	-	-	Autotransporter of unknc	IPR005546:	Yes	IPLSGTQEL	1DAB;Cla	n.d.	132-263	n.d.
YP_0030	Hypothetical	Escherichi	MNKNIRILQ	IVAGNLSN	-	100.0	59	99.6	44	98.7	20	-	-	98.4	21	99.1	26	99.1	26	99.0	24	100.0	45	-	-	82.9	18	98.9	21	99.9	26	99.7	30	-	-	Autotransporter of unknc	IPR005546:	Yes	QTIQLNGN	1DAB;Cla	n.d.	265-383	L1 (160-2
YP_0030	Outer membr	Escherichi	MNKNIRILQ	IVAGNLSN	-	100.0	59	99.6	44	98.7	20	-	-	98.4	21	99.1	26	99.1	26	99.0	24	100.0	45	-	-	83.5	17	98.7	22	99.9	26	99.7	30	-	-	Autotransporter of unknc	IPR005546:	Yes	QTIQLNGN	1DAB;Cla	n.d.	322-432	n.d.
YP_0014	Hypothetical	Escherichi	MNKNIRILQ	IVAGNLSN	-	100.0	59	99.5	44	98.4	20	-	-	98.0	21	98.9	27	98.9	26	98.7	20	100.0	44	-	-	83.1	18	98.9	21	100.0	26	99.7	26	-	-	Autotransporter of unknc	IPR005546:	Yes	QTIQLNGN	1DAB;Cla	n.d.	428-554	n.d.
YP_0033	Adhesin autc	Citrobacter	MNASFKPTI	PAAQTRS	-	100.0	40	99.4	36	98.3	26	-	-	98.3	24	98.8	29	98.7	30	98.8	21	100.0	36	-	-	95.8	29	99.1	24	100.0	20	99.7	32	-	-	Autotransporter of unknc	IPR005546:	Yes	ATPASVIID	1DAB;Cla	IPR011050;IPF	265-383	L2 (358-4
YP_0033	Adhesin autc	Citrobacter	MKIPALSPQ	STLTTLTL	-	100.0	51	99.5	40	98.7	32	-	-	98.6	26	98.9	24	98.9	37	99.0	20	100.0	36	-	-	93.2	29	99.1	25	100.0	29	99.8	36	-	-	Autotransporter of unknc	IPR005546:	Yes	APVKYTLD	1DAB;Cla	IPR011050;IPF	265-383	n.d.
YP_0033	Adhesin autc	Citrobacter	MSRQVSEY	MTNSSSV	-	100.0	127	99.7	36	99.1	29	-	-	98.9	24	99.4	29	99.3	31	99.4	24	100.0	35	-	-	97.1	31	99.3	24	100.0	23	99.8	32	-	-	Autotransporter of unknc	IPR005546:	Yes	EGERIVDS	1DAB;Cla	IPR011050;IPF	265-383	n.d.
YP_0033	Adhesin autc	Citrobacter	MKSQLISFR	TATLSGD	-	100.0	94	48.0	33	-	-	-	-	-	-	-	-	-	-	-	-	100.0	37	-	-	88.4	22	98.6	20	99.9	22	99.6	28	-	-	Autotransporter of unknc	IPR005546:	Yes	AANNGNLL	1DAB;Cla	n.d.	265-383	L3 (174-2
YP_0022	Autotranspor	Salmonella	MPTPQNYS	STLTDLTL	-	100.0	54	99.2	36	97.8	23	-	-	96.9	19	98.4	24	97.9	28	98.0	17	100.0	36	-	-	89.5	20	99.1	22	100.0	25	99.8	29	-	-	Autotransporter of unknc	IPR005546:	Yes	AWVDVDS	1DAB;Cla	IPR011050;IPF	132-263	T3 (270-3
YP_0022	Autotranspor	Salmonella	MPTPQNYS	STLTDLTL	-	100.0	116	99.4	36	98.2	22	-	-	97.7	19	98.8	25	98.5	29	98.5	17	100.0	36	-	-	87.3	20	99.0	22	100.0	25	99.8	29	-	-	Autotransporter of unknc	IPR005546:	Yes	AWVDVDS	1DAB;Cla	IPR011050;IPF	132-263	n.d.
YP_0021	Outer membr	Salmonella	MPTPQNYS	STLTDLTL	-	100.0	116	99.7	36	98.8	20	-	-	98.4	19	99.2	24	99.0	28	99.0	17	100.0	36	-	-	87.3	20	99.0	22	100.0	25	99.8	29	-	-	Autotransporter of unknc	IPR005546:	Yes	AWVDVDS	1DAB;Cla	IPR011050;IPF	132-263	n.d.
NP_4626	Autotranspor	Salmonella	MPTPQNYS	STLTDLTL	-	100.0	54	99.6	36	98.6	22	-	-	98.2	19	99.1	25	98.9	29	98.8	17	100.0	35	-	-	87.3	20	99.1	22	100.0	25	99.8	30	-	-	Autotransporter of unknc	IPR005546:	Yes	AWVDVDS	1DAB;Cla	IPR011050;IPF	n.d.	T3 (270-3
YP_1527	Autotranspor	Salmonella	MPTPQNYS	STLTDLTL	-	100.0	111	99.7	36	98.8	20	-	-	98.4	19	99.2	25	99.0	28	99.0	18	100.0	36	-	-	89.9	20	99.1	22	100.0	25	99.8	29	-	-	Autotransporter of unknc	IPR005546:	Yes	AWVDVDS	1DAB;Cla	IPR011050;IPF	91-203	T3 (270-3
YP_0021	Autotranspor	Salmonella	MPTPQNYS	STLTDLTL	-	100.0	52	99.7	36	98.8	20	-	-	98.4	19	99.2	25	99.0	28	99.0	18	100.0	37	-	-	89.5	18	99.1	22	100.0	24	99.8	29	-	-	Autotransporter of unknc	IPR005546:	Yes	AWVDVDS	1DAB;Cla	IPR011050;IPF	132-263	n.d.
YP_0015	Hypothetical	Salmonella	MPTPQNYS	STLTDLTL	-	100.0	51	99.6	37	98.8	20	-	-	98.4	19	99.1	25	99.0	28	98.9	17	100.0	35	-	-	87.3	20	99.1	22	100.0	25	99.8	30	-	-	Autotransporter of unknc	IPR005546:	Yes	AWVDVDS	1DAB;Cla	IPR011050;IPF	428-554	n.d.
YP_0021	Outer membr	Salmonella	MPTPQNYS	STLTDLTL	-	100.0	55	98.8	36	97.4	23	-	-	96.7	19	98.0	25	97.4	30	97.8	18	100.0	36	-	-	92.4	20	99.0	20	100.0	25	99.7	31	-	-	Autotransporter of unknc	IPR005546:	Yes	AWVDVDS	1DAB;Cla	IPR011050;IPF	132-263	T3 (270-3
YP_2186	Autotranspor	Salmonella	MQEAECQL	STLTDLTL	-	100.0	54	99.5	36	98.5	23	-	-	98.2	19	99.0	24	98.8	28	98.8	17	100.0	35	-	-	87.3	20	99.1	22	100.0	25	99.8	30	-	-	Autotransporter of unknc	IPR005546:	Yes	AWVDVDS	1DAB;Cla	IPR011050;IPF	91-203	T3 (275-3
YP_0026	Autotranspor	Salmonella	MVFPSQAA	STLTDLTL	-	100.0	55	99.8	36	99.2	22	70.4	21	98.9	19	99.5	26	99.3	28	99.3	21	100.0	36	-	-	87.3	20	99.0	22	100.0	25	99.8	29	-	-	Autotransporter of unknc	IPR005546:	Yes	IIPAQASTG	1DAB;Cla	IPR011050;IPF	473-603	n.d.
YP_0033	Adhesin autc	Citrobacter	MKTTYKDLK	STLTDLTL	-	100.0	112	99.0	37	97.9	24	-	-	97.9	23	98.2	30	98.0	36	98.1	19	100.0	38	-	-	96.4	23	99.3	23	100.0	31	99.8	36	-	-	Autotransporter of unknc	IPR005546:	Yes	ELTLDHNN	1DAB;Cla	IPR011050;IPF	265-383	n.d.
YP_0014	Hypothetical	Citrobacter	MKSSRLIQK	SVVSTLTL	-	100.0	64	99.4	35	98.4	25	-	-	98.1	24	98.8	31	98.6	30	98.7	19	100.0	36	-	-	95.8	24	99.2	24	100.0	28	99.7	30	-	-	Autotransporter of unknc	IPR005546:	Yes	AGCSQSSS	1DAB;Cla	IPR011050;IPF	428-554	L2 (176-2
YP_0039	Outer membr	Enterobact	MKQNCHNK	SVVSTLNI	-	100.0	77	99.4	32	98.4	28	-	-	98.1	23	98.9	31	98.8	26	98.8	17	100.0	33	-	-	95.0	24	99.1	22	100.0	28	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	AGCSYTTL	1DAB;Cla	n.d.	91-203	n.d.
YP_0020	Outer membr	Stenotroph	MNASPTRQ	TLDHAGT	-	100.0	60	99.8	36	99.1	15	35.4	21	99.1	16	99.5	34	99.4	31	99.4	25	100.0	36	-	-	92.4	20	99.0	17	100.0	28	99.7	30	-	-	Autotransporter of unknc	IPR005546:	Yes	AELQGPLK	1DAB;Cla	n.d.	132-263	L1 (70-13
YP_0019	Virulence-as	Stenotroph	MVRLPAVA	ALEHAGT	-	100.0	159	99.9	36	99.5	15	78.5	21	99.4	16	99.7	36	99.7	29	99.7	21	100.0	36	-	-	91.4	20	99.0	17	100.0	28	99.8	31	-	-	Autotransporter of unknc	IPR005546:	Yes	ADLTGPLK	1DAB;Cla	n.d.	132-263	L1 (121-1
YP_0020	Outer membr	Stenotroph	MRMMQFTF	TGNLTLG	-	100.0	99	97.5	31	51.2	12	-	-	25.3	14	87.8	25	-	-	72.6	20	100.0	31	-	-	92.3	16	99.0	15	100.0	27	99.7	25	-	-	Autotransporter of unknc	IPR005546:	Yes	VDVGPGQ	1DAB;Cla	n.d.	132-263	L1 (155-2
YP_0019	Autotranspor	Stenotroph	MRMMQFTF	SNTGNLT	-	100.0	92	97.8	31	62.3	17	-	-	51.9	16	90.5	27	37.2	27	80.1	17	100.0	30	-	-	96.0	20	99.2	18	100.0	29	99.8	30	-	-	Autotransporter of unknc	IPR005546:	Yes	ATVNPGQT	1DAB;Cla	n.d.	132-263	n.d.
YP_0013	Autotranspor	Yersinia ps	MKNSNRSP	SNIHHLDI	-	100.0	272	99.4	34	97.4	15	-	-	97.1	15	98.9	21	98.5	29	98.3	21	100.0	34	-	-	86.4	16	99.1	15	100.0	22	99.8	28	-	-	Autotransporter of unknc	IPR005546:	Yes	NPDHEGIV	n.d.	n.d.	428-554	T5 (85-13
YP_0017	Outer membr	Yersinia ps	MKNSNRSP	SNIHHLDI	-	100.0	265	99.2	34	97.2	15	-	-	96.4	15	98.7	21	98.2	29	98.1	21	100.0	34	-	-	86.4	16	99.1	15	100.0	22	99.8	28	-	-	Autotransporter of unknc	IPR005546:	Yes	NPDHEGIV	1DAB;Cla	n.d.	428-554	n.d.
YP_0719	Pertactin fam	Yersinia ps	MKNSNRSP	SNIHHLDI	-	100.0	265	99.5	35	97.9	15	-	-	97.3	15	99.0	21	98.7	29	98.6	22	100.0	34	-	-	87.3	16	99.1	15	100.0	22	99.8	28	-	-	Autotransporter of unknc	IPR005546:	Yes	NPDHEGIV	1DAB;Cla	n.d.	91-203	T5 (85-13
YP_0011	Autotranspor	Yersinia ps	MKNSNRSP	SNIHHLDI	-	100.0	265	98.9	35	94.5	15	-	-	93.4	15	98.0	21	97.5	29	97.2	19	100.0	34	-	-	93.0	18	99.0	16	100.0	22	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	NPDHEGIV	n.d.	n.d.	428-554	T5 (85-13
YP_0018	Outer membr	Yersinia ps	MKNSNRSP	SNIHHLDI	-	100.0	231	99.4	34	97.7	15	-	-	97.0	15	99.0	21	98.6	29	98.6	22	100.0	34	-	-	93.0	18	99.0	16	100.0	22	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	NPDHEGIV	1DAB;Cla	n.d.	132-263	n.d.
YP_0023	Autotranspor	Yersinia ps	MKNSNRSP	SNIHHLDI	-	100.0	262	99.1	35	95.7	15	-	-	93.7	15	98.4	21	97.8	29	97.7	22	100.0	34	-	-	86.4	16	99.1	15	100.0	22	99.8	28	-	-	Autotransporter of unknc	IPR005546:	Yes	NPDHEGIV	1DAB;Cla	n.d.	473-603	n.d.
NP_9942	Autotranspor	Yersinia ps	MKNSNRSP	SNIHHLDI	-	100.0	262	99.3	35	97.4	15	-	-	96.6	15	98.9	21	98.4	29	98.3	19	100.0	34	-	-	86.4	16	99.1	15	100.0	22	99.8	28	-	-	Autotransporter of unknc	IPR005546:	Yes	NLDHEGIV	1DAB;Cla	n.d.	428-554	n.d.
NP_6708	Atp-binding t	Yersinia ps	MKNSNRSP	SNIHHLDI	-	100.0	262	99.1	35	95.7	15	-	-	93.7	15	98.4	21	97.8	29	97.7	22	100.0	34	-	-	86.4	16	99.1	15	100.0	22	99.8	28	-	-	Autotransporter of unknc	IPR005546:	Yes	NPDHEGIV	1DAB;Cla	n.d.	n.d.	n.d.
YP_6531	Putative autc	Yersinia ps	MKNSNRSP	SNIHHLDI	-	100.0	262	99.1	35	95.7	15	-	-	93.7	15	98.4	21	97.8	29	97.7	22	100.0	36	-	-	81.6	20	99.0	15	100.0	20	99.8	27	-	-	Autotransporter of unknc	IPR005546:	Yes	NPDHEGIV	1DAB;Cla	n.d.	180-297	n.d.
YP_6463	Autotranspor	Yersinia ps	MKNSNRSP	SNIHHLDI	-	100.0	262	99.1	35	95.7	15	-	-	93.7	15	98.4	21	97.8	29	97.7	22	100.0	36	-	-	81.6	20	99.0	15	100.0	20	99.8	27	-	-	Autotransporter of unknc	IPR005546:	Yes	NPDHEGIV	1DAB;Cla	n.d.	180-297	n.d.
YP_0035	Putative autc	Yersinia ps	MGKSILNKK	SNIHHLDI	-	100.0	227	98.4	34	50.1	14	-	-	66.7	15	97.4	21	95.3	29	95.1	19	100.0	34	-	-	86.4	16	99.1	15	100.0	22	99.8	28	-	-	Autotransporter of unknc	IPR005546:	Yes	NINTVLSG	1DAB;Cla	n.d.	200-322	n.d.
YP_8368	Outer membr	Burkholder	MAILTGAVP	LTNAGVV	-	100.0	88	40.8	33	-	-	-	-	-	-	-	-	-	-	-	-	100.0	33	-	-	57.8	26	98.6	22	99.9	27	99.7	33	-	-	Autotransporter of unknc	IPR005546:	Yes	ANAPSVVA	1DAB;Cla	n.d.	345-479	TS (430-4
YP_6248	Outer membr	Burkholder	MAILTGAVP	LTNAGVV	-	100.0	88	40.8	33	-	-	-	-	-	-	-	-	-	-	-	-	100.0	31	-	-	-	-	98.6	23	99.9	25	99.5	33	-	-	Autotransporter of unknc	IPR005546:	Yes	ANAPSVVA	1DAB;Cla	n.d.	180-297	TS (430-4
NP_880C	Autotranspor	Bordetella	MRSARRRT	LRNAGEIF	-	100.0	88	81.4	32	-	-	-	-	-	-	-	-	-	-	-	-	100.0	36	-	-	52.5	21	98.9	17	99.9	25	99.6	30	-	-	Autotransporter of unknc	IPR005546:	Yes	ACAPTLAP	1DAB;Cla	n.d.	n.d.	L1 (138-2
NP_8893	Autotranspor	Bordetella	MLAAAGLLC	LRNAGEIF	-	100.0	73	20.7	31	-	-	-	-	-	-	-	-	-	-	-	-	100.0	36	-	-	78.1	20	98.9	16	99.9	25	99.6	27	-	-	Autotransporter of unknc	IPR005546:	Yes	ACAPTLAP	1DAB;Cla	n.d.	428-554	L1 (121-1
YP_0023	Aida autotrar	Escherichi	MNRTYNIVV	NGDVYNS	-	100.0	81	98.5	28	94.8	14	-	-	92.9	14	97.1	27	96.6	27	96.5	23	100.0	34	-	-	57.4	19	98.9	16	100.0	27	99.7	25	-	-	Autotransporter of unknc	IPR005546:	Yes	ESTSEVTN	1DAB;Cla	n.d.	473-603	n.d.
YP_0023	Hypothetical	Escherichi	MKQNKLPG	STVTNLTI	-	100.0	114	98.6	40	94.7	25	-	-	92.5	19	97.0	25	96.4	30	97.5	24	100.0	39	-	-	95.3	18	99.1	20	100.0	27	99.8	32	-	-	Autotransporter of unknc	IPR005546:	Yes	DIVITPSDD	1DAB;Cla	n.d.	473-603	L1 (236-3
YP_0024	Putative oute	Escherichi	MKLKKLPGF	STVSKLTV	-	100.0	63	99.0	39	97.8	21	-	-	97.3	15	98.2	25	97.9	26	98.4	24	100.0	39	-	-	94.0	21	99.1	16	100.0	24	99.7	26	-	-	Autotransporter of unknc	IPR005546:	Yes	EQFIQYDE	1DAB;Cla	n.d.	473-603	n.d.
YP_0017	Autotranspor	Escherichi	MQGNNTIV	TSTVSKLTV	-	100.0	113	98.3	38	95.6	23	-	-	95.0	15	97.4	25	96.3	26	97.5	24	100.0	39	-	-	94.0	21	99.1	16	100.0	24	99.7	26	-	-	Autotransporter of unknc	IPR005546:	Yes	CVSSLDEE	1DAB;Cla	n.d.	428-554	n.d.
YP_0023	Putative oute	Escherichi	MQGKNTIVI	STVSNLTV	-	100.0	111	98.9	39	96.8	23	-	-	96.6	14	97.8	24	97.5	27	98.1	24	100.0	40	-	-	94.8	23	99.1	15	100.0	23	99.7	29	-	-	Autotransporter of unknc	IPR005546:	Yes	CVSSLDEE	1DAB;Cla	n.d.	473-603	L1 (209-3
YP_0032	Aida-i adhesi	Escherichi	MQGKNTIVI	STVSNLTV	-	100.0	112	98.8	39	97.0	23	-	-	97.1	15	97.9	24	97.5	28	98.1	24	100.0	40	-	-	94.8	23	99.1	15	100.0	23	99.7	29	-	-	Autotransporter of unknc	IPR005546:	Yes	CVSSLDEE	1DAB;Cla	n.d.	265-383	n.d.
YP_0022	Putative autc	Escherichi	MQGKNTIVI	STVSNLTV	-	100.0	111	98.3	39	93.4	23	-	-	91.3	14	97.0	24	96.1	26	97.3	24	100.0	41	-	-	94.9	22	99.1	15	100.0	25	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	CVSSLDEE	1DAB;Cla	n.d.	473-603	n.d.
YP_0032	Aida-i adhesi	Escherichi	MQGKNTIVI	STVSNLTV	-	100.0	112	98.4	36	94.4	21	-	-	94.0	14	97.1	24	96.3	28	97.5	24	100.0	40	-	-	94.8	23	99.1	15	100.0	23	99.7	29	-	-	Autotransporter of unknc	IPR005546:	Yes	CVSSLDEE	1DAB;Cla	n.d.	265-383	n.d.
YP_0024	Outer membr	Escherichi	MESLPTGD	TWDIKAD	-	100.0	108	98.7	38	96.3	22	-	-	94.7	14	97.7	24	96.9	26	97.9	24	100.0	38	-	-	97.9	23	99.5	15	100.0	23	99.8	29	-	-	Autotransporter of unknc	IPR005546:	Yes	CVSSLDEE	1DAB;Cla	n.d.	473-603	n.d.
NP_707C	Hypothetical	Shigella fle	MQGKNTIVT	TWDIKAD	-	100.0	113	98.8	38	97.3	23	-	-	96.9	14	97.8	25	97.3	28	98.1	24	100.0	36	-	-	97.8	22	99.4	15	100.0	23	99.8	29	-	-	Autotransporter of unknc	IPR005546:	Yes	CVSSLDEE	1DAB;Cla	n.d.	n.d.	n.d.
NP_8368	Hypothetical	Shigella fle	MQGKNTIVT	STVSNLTV	-	100.0	112	98.0	38	93.7	21	-	-	91.6	14	96.4	25	92.8	27	96.9	24	100.0	39	-	-	95.6	24	99.1	15	100.0	23	99.7	27	-	-	Autotransporter of unknc	IPR005546:	Yes	CVSSLDEE	1DAB;Cla	n.d.	n.d.	n.d.
YP_6886	Hypothetical	Shigella fle	MLSQTSGNI	STVSNLTV	-	100.0	96	97.4	36	38.6	22	-	-	71.2	14	87.8	24	41.3	26	94.8	24	100.0	40	-	-	94.3	23	99.1	15	100.0	25	99.7	27	-	-	Autotransporter of unknc	IPR005546:	Yes	CVSSLDEE	1DAB;Cla	n.d.	180-297	n.d.
YP_0033	Outer membr	Sebaldella	MRKSFYGK	SDLTNFSI	-	100.0	87	99.2	37	97.8	21	-	-	97.6	19	98.4	25	98.1	28	98.4	24	100.0	37	-	-	87.3	28	99.1	20	100.0	25	99.7	29	-	-	Autotransporter of unknc	IPR005546:	Yes	QTIWTEAD	1DAB;Cla	n.d.	265-383	L1 (248-3
YP_0024	Hypothetical	Escherichi	MDSIDVNKT	WNMKDD	-	100.0	81	99.1	37	98.0	22	41.4	17	97.0	19	98.7	25	98.6	27	98.3	27	100.0	36	-	-	97.6	21	99.5	22	100.0	27	99.9	29	-	-	Autotransporter of unknc	IPR005546:	Yes	ATKLVSGT	1DAB;Cla	n.d.	473-603	n.d.
YP_0032	Hypothetical	Cronobact	MQDELVQN	DLTLGGA	-	100.0	208	99.9	27	99.6	17	-	-	99.1	21	99.8	29	99.7	23	99.6	25	100.0	31	-	-	96.5	27	99.5	23	100.0	29	99.7	24	-	-	Self-Associating AutoTra	IPR030930:	Yes	TSINDGGS	1DAB;Cla	n.d.	265-383	L3 (1698-
YP_0014	Hypothetical	Cronobact	MKRSLNCS	NLTLGGS	-	100.0	80	99.8	27	99.4	18	-	-	98.9	24	99.6	31	99.4	28	99.5	26	100.0	30	-	-	95.9	25	99.2	25	100.0	30	99.7	28	-	-	Self-Associating AutoTra	IPR030930:	Yes	FTPDVTSS	1DAB;Cla	n.d.	428-554	L3 (1761-
YP_0041	Outer membr	Pantoea sp	MNKAYRIVV	TLTNAGS	-	100.0	98	99.7	34	99.2	24	-	-	98.8	22	99.5	27	99.3	25	99.3	29	100.0	34	-	-	94.7	26	99.1	21	100.0	30	99.7	31	-	-	Self-Associating AutoTra	IPR030930:	Yes	TYYMGTVV	1DAB;Cla	n.d.	91-203	L1 (647-7
YP_0033	Outer membr	Dickeya da	MFFLPRSSV	TLTNAGA	-	100.0	101	97.6	27	93.7	18	-	-	41.5	17	97.3	26	93.3	22	93.8	22	100.0	31	-	-	94.1	22	99.1	23	100.0	24	99.7	23	-	-	Self-Associating AutoTra	IPR030930:	Yes	AGNTTVPA	1DAB;Cla	n.d.	265-383	L1 (439-5
NP_7087	Outer membr	Shigella fle	MLMKRHLN	DLSHAGC	-	100.0	70	98.7	23	97.3	18	-	-	96.1	20	98.4	25	97.6	24	98.0	24	100.0	27	-	-	90.1	25	99.0	23	100.0	23	99.7	20	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADIVVHPGI	1DAB;Cla	n.d.	n.d.	T4 (145-2
YP_0029	Cp4-44 propl	Escherichi	MKRHLNTC	DLSHAGC	-	100.0	67	98.9	26	97.7	18	-	-	96.5	21	98.5	22	97.9	22	98.1	24	100.0	28	-	-	19.6	27	98.4	22	99.9	19	99.3	22	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADIVVHPGI	1DAB;Cla	n.d.	473-603	T4 (107-1
YP_0261	Cp4-44 propl	Escherichi	MKRHLNTC	DLSHAGC	-	100.0	67	98.9	26	97.7	18	-	-	96.5	21	98.5	22	97.9	22	98.1	24	100.0	27	-	-	88.4	26	98.8	22	99.9	21	99.6	21	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADIVVHPGI	1DAB;Cla	n.d.	91-203	T4 (107-1
YP_0032	Adhesin	Escherichi	MKRHLNTS	DLSHAGC	-	100.0	63	98.7	25	96.9	18	-	-	93.2	19	98.2	23	97.5	20	97.6	22	100.0	27	-	-	61.6	27	98.6	21	99.9	21	99.5	21	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADIVVHPGI	1DAB;Cla	n.d.	265-383	T4 (107-1
YP_0032	Aida-i adhesi	Escherichi	MKRHLNTC	DLSHAGC	-	100.0	67	98.8	25	97.4	18	-	-	96.2	21	98.5	23	97.6	24	97.9	18	100.0	27	-	-	61.6	27	98.6	21	99.9	21	99.5	21	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADIVVHPGI	1DAB;Cla	n.d.	265-383	T4 (107-1
YP_0017	Cp4-44 propl	Escherichi	MKRHLNTC	DLSHAGC	-	100.0	67	98.9	26	97.7	18	-	-	96.5	21	98.5	22	97.9	22	98.1	24	100.0	28	-	-	19.6	27	98.4	22	99.9	19	99.3	22	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADIVVHPGI	1DAB;Cla	n.d.	428-554	T4 (107-1
YP_0035	Type v secre	Escherichi	MKRHLNTS	YFDNQTT	-	100.0	69	99.3	24	97.9	19	-	-	97.7	17	98.9	23	98.6	22	98.7	21	100.0	23	-	-	70.3	25	98.9	18	99.9	20	99.6	23	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADTVVQAG	1DAB;Cla	n.d.	200-322	T1 (425-4
YP_0030	Calcium-bind	Escherichi	MKRHLNTS	AALSRAV	-	100.0	67	99.3	25	98.3	20	-	-	97.8	16	98.9	19	98.6	25	98.7	25	100.0	25	-	-	21.5	27	98.7	16	99.9	17	99.5	25	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADKVVQAG	1DAB;Cla	n.d.	265-383	T4 (143-2
YP_0022	Pertactin fam	Escherichi	MKRHLNTS	AALSRAV	-	100.0	67	99.3	25	98.3	20	-	-	97.8	16	98.9	19	98.6	25	98.7	25	100.0	25	-	-	21.5	27	98.7	16	99.9	17	99.5	25	-	-	Self-Associating AutoTra	IPR030930:	Yes	ADKVVQAG	1DAB;Cla	n.d.	473-603	T1 (426-4
YP_0020	Autoporter pr	Cupriavidu	MLLKTQNKI	NAGTIDLT	-	100.0	192	95.4	32	-	-	-	-	-	-	58.3	23	-	-	-	-	100.0	32	-	-	65.7	18	98.6	18	99.9	24	99.7	32	-	-	Autotransporter of unknc	IPR005546:	Yes	NTLTNAGT	1DAB;Cla	n.d.	132-263	L1 (140-2
YP_0020	Outer membr	Stenotroph	MSRAVPRA	NAGRIDLT	-	100.0	158	96.6	37	20.8	16	-	-	-	-	87.3	24	41.6	22	34.3	22	100.0	38	-	-	42.9	21	98.9	17	99.9	28	99.5	27	-	-	Autotransporter of unknc	IPR005546:	Yes	GCNNPALT	n.d.	n.d.	132-263	L1 (136-2
YP_0019	Autotranspor	Stenotroph	MYRAVPRA	NAGRIDLT	-	100.0	158	97.1	38	25.1	18	-	-	-	-	89.0	24	45.6	23	44.8	21	100.0	38	-	-	-	-	98.6	14	99.9	29	99.5	28	-	-	Autotransporter of unknc	IPR005546:	Yes	GCNSPAPT	n.d.	n.d.	132-263	n.d.
YP_0025	Rhizobiocin/r	Agrobacter	MQYRIAAKS	NAGRIDLT	-	100.0	170	82.5	29	-	-	-	-	-	-	-	-	-	-	-	-	100.0	39	-	-	88.9	27	98.8	17	99.9	27	99.7	30	-	-	Autotransporter of unknc	IPR005546:	Yes	ACTFAAPT	1DAB;Cla	n.d.	473-603	L1 (214-2
YP_0013	Outer membr	Ochrobact	MHCRNIGSF	NAGRIDLT	-	100.0	155	72.7	37	-	-	-	-	-	-	-	-	-	-	-	-	100.0	39	-	-	78.3	21	98.8	18	99.9	28	99.6	29	-	-	Autotransporter of unknc	IPR005546:	Yes	ACSFSPTL	1DAB;Cla	n.d.	428-554	L3 (306-3
NP_3962	Autotranspor	Agrobacter	MRHRATDC	NSGRIDLT	-	100.0	167	96.4	35	-	-	-	-	-	-	67.5	26	-	-	-	-	100.0	37	-	-	54.9	26	99.0	20	99.9	26	99.6	30	-	-	Autotransporter of unknc	IPR005546:	Yes	ACNFPAPV	1DAB;Cla	n.d.	n.d.	L3 (286-3
YP_0041	Outer membr	Mesorhizol	MRHVKNSLI	NAGRIDLT	-	100.0	176	98.4	32	93.6	18	-	-	90.4	15	96.7	24	94.1	31	92.1	23	100.0	31	-	-	12.5	22	98.9	15	100.0	25	99.6	30	-	-	Autotransporter of unknc	IPR005546:	Yes	ACVLAPSA	1DAB;Cla	n.d.	91-203	L1 (275-3
YP_2355	Outer membr	Pseudomo	MKIRCVHLY	NAGTLDLS	-	100.0	170	98.6	33	95.8	17	-	-	90.8	21	97.4	25	96.3	27	95.1	21	100.0	35	-	-	36.5	15	98.9	21	99.9	24	99.5	29	-	-	Autotransporter of unknc	IPR005546:	Yes	GVQVCDS	1DAB;Cla	n.d.	180-297	n.d.
NP_7925	Autotranspor	Pseudomo	MRRHLRGS	NAGTLDLT	-	100.0	172	97.3	32	47.7	16	-	-	-	-	89.5	24	-	-	22.4	20	100.0	34	-	-	58.3	17	98.9	21	100.0	21	99.6	26	-	-	Autotransporter of unknc	IPR005546:	Yes	ACNLAPTP	1DAB;Cla	n.d.	n.d.	n.d.
YP_0012	Outer membr	Pseudomo	MRLRLMLTL	NAGTLDLT	-	100.0	89	97.9	32	88.9	16	-	-	81.5	16	97.1	22	94.2	21	92.7	18	100.0	33	-	-	31.4	15	98.9	15	100.0	26	99.6	25	-	-	Autotransporter of unknc	IPR005546:	Yes	NCTFAPGP	1DAB;Cla	n.d.	428-554	L1 (245-3
YP_2980	Outer membr	Ralstonia	MPHRSSRP	NAGVIDLT	-	100.0	93	97.9	35	83.5	16	-	-	69.8	18	95.3	21	92.9	23	88.2	15	100.0	38	-	-	83.8	15	98.9	18	100.0	24	99.8	26	-	-	Autotransporter of unknc	IPR005546:	Yes	ACDLAAPA	1DAB;Cla	n.d.	180-297	n.d.
YP_5564	Outer membr	Burkholder	MLHKGRLS	LTNSASNI	-	100.0	214	94.6	35	-	-	-	-	-	-	20.9	29	-	-	23.6	24	100.0	31	-	-	97.3	26	99.3	23	100.0	29	99.8	29	-	-	Autotransporter of unknc	IPR005546:	Yes	ACAPDGQ	n.d.	n.d.	180-297	L1 (328-3
YP_0036	Outer membr	Burkholder	MKMRAIVR	LTNNASQ	-	100.0	177	99.0	29	98.0	20	-	-	97.2	17	98.4	38	98.5	25	98.1	26	100.0	29	-	-	98.2	27	99.5	20	100.0	37	99.8	26	-	-	Autotransporter of unknc	IPR005546:	Yes	LTINGPSAF	n.d.	n.d.	200-322	L1 (191-2
YP_0011	Outer membr	Burkholder	MNHSFRSIV	LNNAGLV	-	100.0	109	99.4	37	98.8	21	-	-	98.4	21	99.2	33	98.7	26	98.9	22	100.0	36	-	-	87.4	26	98.9	21	99.9	33	99.6	24	-	-	Autotransporter of unknc	IPR005546:	Yes	STCGNGSA	1DAB;Cla	n.d.	428-554	L1 (1081-
YP_0011	Outer membr	Enterobact	MKSRNKCV	MTADSNV	-	100.0	107	99.7	40	98.7	23	-	-	98.7	20	99.2	36	99.2	30	99.1	27	100.0	39	-	-	97.5	20	99.4	23	100.0	31	99.8	31	-	-	Autotransporter of unknc	IPR005546:	Yes	SECMTANC	1DAB;Cla	n.d.	428-554	L1 (999-1
YP_0011	Outer membr	Enterobact	MNEVNYSKI	NASSTVN	-	100.0	90	98.8	42	95.5	25	-	-	96.5	18	97.3	24	97.1	28	96.9	25	100.0	40	-	-	95.3	20	99.2	21	100.0	27	99.8	30	-	-	Autotransporter of unknc	IPR005546:	Yes	KSSSCNTS	1DAB;Cla	n.d.	428-554	n.d.
YP_7675	Adhesin autc	Rhizobium	MRIYRWLS	LNDPSLIC	-	100.0	91	98.9	24	98.3	22	-	-	97.7	22	98.5	28	98.4	26	98.2	26	100.0	29	-	-	97.4	23	99.3	21	100.0	27	99.7	25	-	-	Autotransporter of unknc	IPR005546:	Yes	AVPGLDAR	1DAB;Cla	n.d.	180-297	L1 (155-2
YP_2979	Outer membr	Ralstonia	MQKNKTKL	VWNMTGI	-	100.0	165	99.2	33	98.1	15	-	-	97.5	21	98.7	36	98.5	26	98.5	25	100.0	31	-	-	98.4	26	99.5	23	100.0	35	99.8	28	-	-	Autotransporter of unknc	IPR005546:	Yes	TTITMTPA	1DAB;Cla	n.d.	180-297	L1 (354-4
YP_0013	Autotranspor	Janthinoba	MIKVMHRFI	TQTLTNA	-	100.0	61	99.2	33	98.0	18	-	-	98.1	17	98.8	27	98.8	23	98.6	23	100.0	33	-	-	61.4	24	99.1	18	100.0	27	99.7	25	-	-	Autotransporter of unknc	IPR005546:	Yes	QTIIGPSPP	1DAB;Cla	n.d.	428-554	L1 (254-3
YP_0017	Outer membr	Yersinia ps	MKKNRSTLS	DADVESM	-	100.0	63	99.0	28	98.1	18	24.2	20	97.7	18	98.4	30	98.2	29	98.2	25	100.0	29	-	-	96.7	18	98.9	16	99.9	29	99.7	28	-	-	Autotransporter of unknc	IPR005546:	Yes	VQTASITD	1DAB;Cla	n.d.	428-554	L1 (161-2
YP_0723	Pertactin fam	Yersinia ps	MKQNRSTLS	DADVESM	-	100.0	63	99.0	27	98.1	20	-	-	97.6	17	98.5	29	98.3	25	98.3	21	100.0	29	-	-	96.8	18	98.8	16	99.9	31	99.7	27	-	-	Autotransporter of unknc	IPR005546:	Yes	VQTASITD	1DAB;Cla	n.d.	91-203	n.d.
YP_6495	Putative autc	Yersinia pe	MKTNRSTLS	DADVESM	-	100.0	63	98.3	29	96.7	18	-	-	95.3	17	97.3	30	96.5	28	96.5	21	100.0	29	-	-	96.8	18	98.8	16	99.9	31	99.7	27	-	-	Autotransporter of unknc	IPR005546:	Yes	VQTASITD	1DAB;Cla	n.d.	180-297	n.d.
YP_6537	Putative autc	Yersinia pe	MKTNRSTLS	HVLGQMN	-	100.0	63	98.3	29	96.7	18	-	-	95.3	17	97.3	30	96.5	28	96.5	21	100.0	31	-	-	95.7	21	98.7	18	99.9	27	99.7	32	-	-	Autotransporter of unknc	IPR005546:	Yes	VQTASITD	1DAB;Cla	n.d.	180-297	L1 (161-2
YP_0035	Putative autc	Yersinia pe	MKTNRSTLS	HVLGQMN	-	100.0	63	98.3	29	96.7	18	-	-	95.3	17	97.3	30	96.5	28	96.5	21	100.0	31	-	-	95.7	21	98.7	18	99.9	27	99.7	32	-	-	Autotransporter of unknc	IPR005546:	Yes	VQTASITD	1DAB;Cla	n.d.	200-322	n.d.
YP_0016	Putative autc	Yersinia pe	MKTNRSTLS	HVLGQMN	-	100.0	63	98.3	29	96.7	18	-	-	95.3	17	97.3	30	96.5	28	96.5	21	100.0	31	-	-	95.7	21	98.7	18	99.9	27	99.7	32	-	-	Autotransporter of unknc	IPR005546:	Yes	VQTASITD	1DAB;Cla	n.d.	428-554	n.d.
YP_0011	Putative autc	Yersinia pe	MKTNRSTLS	HVLGQMN	-	100.0	63	98.3	29	96.7	18	-	-	95.3	17	97.3	30	96.5	28	96.5	21	100.0	31	-	-	95.7	21	98.7	18	99.9	27	99.7	32	-	-	Autotransporter of unknc	IPR005546:	Yes	VQTASITD	1DAB;Cla	n.d.	428-554	n.d.
YP_0023	Autotranspor	Yersinia pe	MKTNRSTLS	HVLGQMN	-	100.0	63	98.3	29	96.7	18	-	-	95.3	17	97.3	30	96.5	28	96.5	21	100.0	31	-	-	95.7	21	98.7	18	99.9	27	99.7	32	-	-	Autotransporter of unknc	IPR005546:	Yes	VQTASITD	1DAB;Cla	n.d.	473-603	n.d.
YP_0018	Outer membr	Yersinia ps	MKKNRSTLS	HVLGQMN	-	100.0	60	98.1	29	96.1	21	-	-	95.4	17	97.1	30	97.0	31	96.3	28	100.0	30	-	-	96.2	21	98.7	18	99.9	31	99.7	32	-	-	Autotransporter of unknc	IPR005546:	Yes	VQTASITD	1DAB;Cla	n.d.	132-263	L1 (161-2
YP_0014	Autotranspor	Yersinia ps	MKTNRSTLS	HVLGQMN	-	100.0	59	99.1	29	98.1	18	-	-	97.6	18	98.7	25	98.4	29	98.5	23	100.0	29	-	-	97.7	19	99.1	20	100.0	31	99.8	29	-	-	Autotransporter of unknc	IPR005546:	Yes	VQTASITD	1DAB;Cla	n.d.	428-554	L1 (161-2
YP_0010	Autotranspor	Yersinia er	MNINNIARLI	SMHILGQI	-	100.0	70	98.8	28	97.4	21	-	-	96.7	18	97.7	28	97.7	31	97.5	27	100.0	34	-	-	96.6	21	98.5	19	99.9	30	99.7	29	-	-	Autotransporter of unknc	IPR005546:	Yes	AQTVSVTD	1DAB;Cla	n.d.	428-554	L1 (159-2
YP_0042	Outer membr	Rahnella s	MEEYREHPI	LTGDSTIS	-	100.0	95	97.4	24	93.3	22	-	-	93.3	17	96.8	25	95.9	28	93.0	30	100.0	25	-	-	97.7	19	99.3	22	100.0	25	99.9	27	-	-	Autotransporter of unknc	IPR005546:	Yes	TTTFSTGIT	1DAB;Cla	n.d.	91-203	L1 (172-2
NP_8830	Autotranspor	Bordetella	MLRHLGWL	MTGSSTIT	-	100.0	54	99.8	32	99.3	23	42.5	22	99.1	19	99.7	35	99.5	27	99.6	28	100.0	30	-	-	97.1	27	99.4	19	100.0	36	99.8	31	-	-	Autotransporter of unknc	IPR005546:	Yes	ADYPGLTL	1DAB;Cla	n.d.	n.d.	L1 (366-4
NP_8873	Autotranspor	Bordetella	MLRHLGWL	STITNLVN	-	100.0	55	99.5	31	98.9	23	-	-	98.5	14	99.4	34	99.2	27	99.3	28	100.0	31	-	-	96.2	25	99.2	18	100.0	36	99.8	29	-	-	Autotransporter of unknc	IPR005546:	Yes	ADYPGLTL	1DAB;Cla	n.d.	n.d.	L1 (366-4
YP_2961	Outer membr	Ralstonia	MLARSSLAL	GAIVFGD	-	100.0	25	98.5	36.0	94.3	16	-	-	92.9	15	94.3	33	86.3	27	93.3	22	100.0	36	-	-	92.7	21	98.3	15	99.9	36	99.5	24	-	-	Autotransporter of unknc	IPR005546:	Yes	AECSPWQ	1DAB;Cla	n.d.	180-297	L1 (269-3
YP_7863	Autotranspor	Bordetella	VSSYASGD	TELTTSSS	-	100.0	184	96.2	32	31.2	19	-	-	42.9	16	93.6	29	88.5	27	94.3	25	100.0	33	-	-	96.7	20	99.3	19	100.0	29	99.8	25	-	-	Autotransporter of unknc	IPR005546:	Yes	DSKLVSER	1DAB;Cla	n.d.	180-297	n.d.
YP_7873	Autotranspor	Bordetella	MRRARPSN	SRITRLNN	-	100.0	74	98.0	32	92.6	22	-	-	94.3	19	97.3	31	96.0	24	96.5	20	100.0	34	-	-	95.4	22	99.2	23	100.0	25	99.7	25	-	-	Autotransporter of unknc	IPR005546:	Yes	AQVYANGT	1DAB;Cla	n.d.	180-297	L1 (151-2
YP_0033	Outer membr	Veillonella	MKLQRNVS	LYKGSKV	-	93.2	55	90.3	7.0	-	-	-	-	-	-	93.2	9	46.8	8	33.8	16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Autotransporter of unknc	IPR005546:	Yes	TEINATSDK	n.d.	n.d.	265-383	L2 (140-2
YP_0039	Serine protea	Pantoea va	MFDDDWNT	NSNVSSLI	-	100.0	229	95.5	15	96.7	30	63.1	33	93.2	30	95.1	31	92.9	15	72.1	17	99.3	14	-	-	99.8	32	100.0	30	99.5	33	98.0	18	-	-	Protease / Peptidase S6	IPR005546:	Yes	NWTMTGN	3ML3;Clas	n.d.	200-322	n.d.
YP_5393	Hypothetical	Escherichi	MKLVTRMEI	WNVTGSS	-	98.0	57	98.0	21.0	-	-	-	-	73.6	14	87.2	27	75.5	20	94.8	18	85.0	32	-	-	-	-	-	-	18.4	27	-	-	-	-	Autotransporter of unknc	IPR005546:	Yes	TDSTVSTD	1DAB;Cla	n.d.	180-297	L1 (104-1
NP_7523	Hypothetical	Escherichi	MKLVTRMEI	WNVTGSS	-	96.3	55	80.3	20	-	-	-	-	-	-	-	-	-	-	-	-	82.7	32	-	-	3.6	18	21.9	23	96.3	26	39.6	21	-	-	Autotransporter of unknc	IPR005546:	Yes	TDSTVSTD	1DAB;Cla	n.d.	n.d.	L1 (104-1
YP_0023	Adhesin; oute	Escherichi	MKLVTRMEI	WNVTGSS	-	98.0	57	98.0	21.0	-	-	-	-	73.6	14	87.2	27	75.5	20	94.8	18	85.1	32	-	-	-	-	-	-	26.0	29	-	-	-	-	Autotransporter of unknc	IPR005546:	Yes	TDSTVSTD	1DAB;Cla	n.d.	473-603	n.d.
YP_8515	Hypothetical	Escherichi	MENFFMKN	WNVTGSS	-	94.2	49	59.3	20	-	-	-	-	-	-	-	-	-	-	-	-	83.1	32	-	-	-	-	12.7	23	94.2	27	25.9	22	-	-	Autotransporter of unknc	IPR005546:	Yes	TDSTVSTD	1DAB;Cla	n.d.	345-479	L1 (98-17
YP_0014	Outer membr	Escherichi	MKNSKAFYI	WNVTGSS	-	98.7	45	81.1	25	-	-	-	-	-	-	-	-	-	-	-	-	98.6	28	-	-	-	-	-	-	98.7	20	-	-	96.0	15	Autotransporter of unknc	IPR005546:	Yes	ADNAVSTT	1DAB;Cla	n.d.	428-554	L1 (93-16
YP_0022	Putative autc	Escherichi	MKNSKAFYI	WNVTGSS	-	98.1	45	96.2	26.0	-	-	-	-	-	-	-	-	-	-	-	-	95.0	33		-	6.1	22	54.6	22	98.1	26	94.1	23	97.0	21.0	Autotransporter of unknc	IPR005546:	Yes	ADNAVSTT	1DAB;Cla	n.d.	473-603	n.d.

YP_0023	Putative adh	Escherichi	MINATDSTI	WNVTGSS	-	100.0	178	87.8	25	-	-	-	-	-	-	25.8	16	37.5	15	44.9	15	100.0	26	-	-	98.7	15	99.2	19	100.0	23	99.8	15	99.9	21	Autotransporter of unknc	IPR005546:	Yes	LNLSVADS	1DAB;Cla	n.d.	473-603	n.d.
YP_0041	Hypothetical	Bartonella	MGTVIDVDI	SSISTMIL	-	100.0	135	99.2	26	98.2	20	-	-	97.2	18	98.9	26	98.6	20	98.5	24	100.0	27	-	-	92.1	18	99.0	20	100.0	19	99.7	20	-	-	Autotransporter of unknc	IPR005546:	Yes	EADSIIYITS	1DAB;Cla	IPR011050	91-203	n.d.
YP_7194	Autotranspor	Haemophil	MPIAVLLQLI	DGTLNNG	-	99.9	208	98.9	26	96.9	13	-	-	96.7	14	97.7	17	97.1	19	97.2	15	99.9	29	-	-	-	-	98.3	9	99.8	17	98.6	23	-	-	Autotransporter of unknc	IPR005546:	Yes	VTDLVYKN	1DAB;Cla	IPR011050	180-297	n.d.
YP_0040	Integral mem	Neisseria	MKPFPYRK	ALSGTSY	-	100.0	70	99.1	23	97.5	13	-	-	97.7	11	97.6	15	97.5	19	97.2	18	100.0	25	-	-	-	-	98.5	14	99.9	14	99.2	17	-	-	Autotransporter of unknc	IPR005546:	Yes	AEYSSINE	n.d.	IPR011050	91-203	n.d.
YP_0026	Autotranspor	Salmonella	MHSWKKKL	GVVAMNS	-	97.5	40	72.4	15	-	-	-	-	-	-	76.2	17	31.8	15	54.6	16	97.5	26	-	-	-	-	-	-	96.1	38	76.4	29	-	-	Autotransporter of unknc	IPR005546:	Yes	TTYDTFGY	n.d.	IPR011050;IPF	473-603	n.d.
YP_1515	Flagellin stru	Salmonella	MHSWKKKL	GVIAMTSI	-	97.7	40	85.1	20	-	-	-	-	-	-	-	-	-	-	-	-	97.7	26		-		-	-	-	95.9	31	67.6	21	81.1	30.0	Autotransporter of unknc	IPR005546:	Yes	TTYNTFGY	1DAB;Cla	IPR011050;IPF	91-203	T8 (450-4
YP_0015	Hypothetical	Salmonella	MHSWKKKL	GVIAMTSI	-	97.7	40	96.2	17	79.0	8	-	-	38.2	7	85.7	13	83.5	16	85.0	17	97.7	26	-	-	-	-	-	-	95.9	31	73.4	21	-	-	Autotransporter of unknc	IPR005546:	Yes	TTYNTFGY	1DAB;Cla	IPR011050;IPF	428-554	T8 (461-5
YP_0023	Putative flag	Escherichi	MHSWKKKL	GLIAMNSI	-	97.3	39	28.2	14	-	-	-	-	-	-	-	-	-	-	-	-	97.3	27	-	-	-	-	-	-	95.6	33	82.7	21	-	-	Autotransporter of unknc	IPR005546:	Yes	ATNDISGQ	n.d.	IPR011050;IPF	473-603	L1 (445-5
YP_0017	Outer membr	Escherichi	MHSWKKKL	GLIAMNSI	-	97.4	40	78.0	16	-	-	-	-	-	-	-	-	-	-	-	-	97.4	27	-	-	-	-	-	-	95.8	33	84.2	21	-	-	Autotransporter of unknc	IPR005546:	Yes	ATNDISGQ	n.d.	IPR011050;IPF	428-554	L1 (453-5
YP_0042	Outer membr	Rahnella s	MHAWNKKF	VLTDNGT	-	98.1	38	95.7	16	28.2	8	-	-	-	-	93.8	17	82.5	15	92.9	16	98.1	22	-	-	-	-	-	-	97.7	30	92.9	23	-	-	Autotransporter of unknc	IPR005546:	Yes	AISADISGT	1DAB;Cla	IPR011050;IPF	91-203	n.d.
YP_0036	Hypothetical	Enterobact	MQTWKKKL	LGQVDTH	-	98.7	44	98.7	17	74.5	15	-	-	61.5	10	94.9	15	95.2	17	94.9	19	98.0	23	-	-	-	-	-	-	97.4	29	92.5	22	-	-	Autotransporter of unknc	IPR005546:	Yes	KDISGTTYN	1DAB;Cla	IPR011050;IPF	200-322	n.d.
YP_0014	Hypothetical	Cronobact	MHSWKKKL	LGQVDAD	-	98.0	41	98.0	14	50.0	15	-	-	23.5	6	92.4	16	86.2	10	91.0	19	98.0	25	-	-	-	-	-	-	97.5	29	94.5	21	-	-	Autotransporter of unknc	IPR005546:	Yes	ATDISGKSY	1DAB;Cla	IPR011050;IPF	428-554	L1 (425-5
YP_8515	Structural prc	Escherichi	MHSWKKKL	GVVAMNS	-	97.6	40	66.9	19	-	-	-	-	-	-	-	-	-	-	-	-	97.6	26	-	-	-	-	-	-	95.8	33	79.7	27	-	-	Autotransporter of unknc	IPR005546:	Yes	ANYDTWG	n.d.	n.d.	345-479	n.d.
NP_7524	Structural prc	Escherichi	MHSWKKKL	GVVAMNS	-	97.6	40	21.3	14	-	-	-	-	-	-	-	-	-	-	23.4	16	97.6	26	-	-	-	-	-	-	95.7	33	78.8	27	-	-	Autotransporter of unknc	IPR005546:	Yes	ANYDTWG	1DAB;Cla	n.d.	n.d.	T3 (479-5
YP_0030	Hypothetical	Escherichi	MHSWKKKL	SSDYVLN	-	97.6	41	87.6	19	-	-	-	-	-	-	38.9	16	-	-	40.8	15	97.6	27		-	-	-	-	-	96.4	40	76.5	28	89.7	28.0	Autotransporter of unknc	IPR005546:	Yes	ANYDTWTY	1DAB;Cla	n.d.	322-432	n.d.
YP_0022	Putative auto	Escherichi	MHSWKKKL	SSDYVLN	-	97.5	41	96.2	17.0	-	-	-	-	-	-	38.0	21	-	-	57.9	16	97.5	27	-	-	-	-	-	-	95.3	32	74.4	28	-	-	Autotransporter of unknc	IPR005546:	Yes	ANYDTWTY	1DAB;Cla	n.d.	473-603	L1 (425-5
YP_0017	Outer membr	Escherichi	MLESSGYGI	GIVAMNSI	-	97.6	75	50.7	20	-	-	-	-	-	-	-	-	-	-	-	-	97.6	27	-	-	-	-	-	-	96.6	40	76.1	28	-	-	Autotransporter of unknc	IPR005546:	Yes	ANFDPSGH	n.d.	n.d.	428-554	n.d.
YP_6896	Abc transpor	Shigella fle	MHQSGSVS	LAMSGAN	-	100.0	45	39.0	19	-	-	-	-	-	-	85.1	38	63.4	30	41.6	32	99.9	23	-	-	88.2	27	98.4	23	100.0	40	99.8	30	-	-	Self-Associating AutoTra	IPR030930:	Yes	STVEYGET	1DAB;Cla	n.d.	180-297	T4 (51-10
YP_0024	Putative auto	Escherichi	MASTVEYGI	LAMSDAN	-	100.0	162	79.6	17	-	-	-	-	-	-	94.2	39	84.7	30	90.5	32	99.9	23	-	-	94.9	27	98.6	21	100.0	38	99.7	27	-	-	Self-Associating AutoTra	IPR030930:	Yes	AIKTTTRAN	1DAB;Cla	n.d.	473-603	T4 (168-2
YP_0024	Putative auto	Escherichi	MASTVEYGI	LAMSGAN	-	100.0	162	54.7	17	-	-	-	-	-	-	93.4	44	85.9	27	82.3	36	99.9	23	-	-	93.8	29	98.7	26	100.0	42	99.8	35	-	-	Self-Associating AutoTra	IPR030930:	Yes	AIKTTTRAN	1DAB;Cla	n.d.	473-603	T4 (168-2
YP_0032	Autotranspor	Escherichi	MASTVEYGI	LAMSDAN	-	100.0	164	25.2	16	-	-	-	-	-	-	90.9	38	53.9	28	72.0	31	99.9	22	-	-	95.8	27	98.8	25	100.0	38	99.8	28	-	-	Self-Associating AutoTra	IPR030930:	Yes	AIKTTTRAN	1DAB;Cla	n.d.	265-383	T4 (168-2
YP_0017	Autotranspor	Escherichi	MASTVEYGI	MMDEPVT	-	100.0	162	87.1	17	-	-	-	-	-	-	96.3	38	93.2	28	94.6	33	99.9	22	-	-	95.8	27	98.8	25	100.0	38	99.8	28	99.9	29.0	Self-Associating AutoTra	IPR030930:	Yes	AIKTTTRAN	1DAB;Cla	n.d.	428-554	T4 (168-2
YP_0030	Pertactin	Escherichi	MHQSGSVS	MMDEPVT	-	100.0	45	-	-	-	-	-	-	-	-	65.0	38	-	-	-	-	99.9	22	-	-	92.2	26	98.4	26	100.0	39	99.8	30	-	-	Self-Associating AutoTra	IPR030930:	Yes	STVEYGET	1DAB;Cla	n.d.	322-432	T4 (51-10
YP_0014	Autotranspor	Escherichi	MHQSGSVS	MMDEPVT	-	100.0	45	-	-	-	-	-	-	-	-	65.0	38	-	-	-	-	99.9	22	-	-	95.8	27	98.8	25	100.0	38	99.8	28	99.9	29.0	Self-Associating AutoTra	IPR030930:	Yes	STVEYGET	1DAB;Cla	n.d.	428-554	T4 (51-10
YP_0017	Outer membr	Escherichi	MHQSGSVS	MMDEPVT	-	100.0	45	70.6	16	-	-	-	-	-	-	82.4	38	28.0	30	36.1	30	99.9	22	-	-	90.5	30	98.5	23	100.0	40	99.8	30	-	-	Self-Associating AutoTra	IPR030930:	Yes	STVEYGET	1DAB;Cla	n.d.	428-554	n.d.
YP_0022	Putative auto	Escherichi	MASTVEYGI	MMDEPVT	-	100.0	160	36.6	16	-	-	-	-	-	-	89.4	38	47.5	28	56.5	30	99.9	22	-	-	90.5	30	98.5	23	100.0	40	99.8	30	-	-	Self-Associating AutoTra	IPR030930:	Yes	AIKASTRVN	1DAB;Cla	n.d.	473-603	n.d.
YP_0035	Autotranspor	Escherichi	MHQSGSVS	MMDEPVT	-	100.0	45	-	-	-	-	-	-	-	-	26.5	38	-	-	-	-	99.9	21	-	-	91.8	26	98.5	22	100.0	39	99.8	30	-	-	Self-Associating AutoTra	IPR030930:	Yes	STVEYGET	1DAB;Cla	n.d.	200-322	T4 (51-10
YP_4085	Abc transpor	Shigella bo	MNGTAEYS	MMDEPVT	-	100.0	192	59.0	17	-	-	-	-	-	-	78.6	38	-	-	-	-	99.9	22	-	-	90.5	30	98.5	23	100.0	40	99.8	30	-	-	Self-Associating AutoTra	IPR030930:	Yes	GQFSIKDG	1DAB;Cla	n.d.	180-297	n.d.
YP_4025	Atpase	Shigella dy	MEVFGTNRI	MMDEPVT	-	100.0	96	70.7	16	-	-	-	-	-	-	89.5	38	43.2	30	41.0	32	99.9	22	-	-	89.0	30	98.4	23	100.0	40	99.8	30	-	-	Self-Associating AutoTra	IPR030930:	Yes	GQFSIKDG	1DAB;Cla	n.d.	180-297	n.d.
YP_0011	Autotranspor	Yersinia pe	MKNSNTLN	AHFALND	-	99.9	44	97.3	15	88.6	19	-	-	93.5	15	97.6	28	97.7	25	98.0	16	99.9	21	-	-	95.1	16	98.8	15	99.9	30	99.8	26	-	-	Self-Associating AutoTra	IPR030930:	Yes	VSQIATTD	1DAB;Cla	n.d.	428-554	n.d.
YP_0016	Pertactin fam	Yersinia pe	MKNSNTLN	AHFALND	-	99.9	44	97.6	15	93.1	18	-	-	94.2	17	97.8	32	98.2	27	98.4	21	99.9	21	-	-	95.1	16	98.8	15	99.9	30	99.8	26	-	-	Self-Associating AutoTra	IPR030930:	Yes	VSQIATTD	1DAB;Cla	n.d.	428-554	T4 (545-6
YP_6471	Autotranspor	Yersinia pe	MKNSNTLN	AHFALND	-	99.9	48	98.1	14.0	73.3	18	-	-	82.5	15	97.2	34	97.6	27	97.8	16	99.9	23	-	-	95.7	17	98.8	15	99.9	31	99.8	28	-	-	Self-Associating AutoTra	IPR030930:	Yes	VSQIATTD	n.d.	n.d.	180-297	n.d.
YP_0023	Autotranspor	Yersinia pe	MSQIATTDT	AHFALND	-	99.9	164	97.7	21	94.4	18	-	-	95.9	15	98.1	31	98.3	27	98.5	20	99.9	22	-	-	95.7	16	98.7	14	99.9	30	99.8	27	-	-	Self-Associating AutoTra	IPR030930:	Yes	AEDALYLK	1DAB;Cla	n.d.	473-603	T4 (516-5
YP_0035	Putative auto	Yersinia pe	MIAETTPEA	AHFALND	-	100.0	122	97.5	13.0	88.3	18	-	-	89.5	17	97.3	30	97.6	27	98.1	21	100.0	22	-	-	95.1	17	98.7	15	100.0	31	99.8	26	-	-	Self-Associating AutoTra	IPR030930:	Yes	AEDALYLK	1CII;Class	n.d.	200-322	n.d.
YP_0018	Outer membr	Yersinia ps	MGEITQEQ	AHFALND	-	100.0	116	97.2	18	91.2	19	-	-	94.8	15	97.8	28	97.9	26	98.2	21	99.9	21	-	-	96.3	15	98.8	15	100.0	30	99.8	26	-	-	Self-Associating AutoTra	IPR030930:	Yes	KDIADQVTI	1DAB;Cla	n.d.	428-554	T4 (410-4
YP_0033	Putative auto	Citrobacter	MNKVYNTV	GDATVGA	-	100.0	88	98.5	20	97.1	19	-	-	97.0	15	98.5	42	98.3	36	98.7	29	99.9	22	-	-	93.7	27	98.8	17	100.0	41	99.8	34	-	-	Self-Associating AutoTra	IPR030930:	Yes	VDYNNETL	1DAB;Cla	n.d.	265-383	L1 (921-9
NP_4171	Adhesin-like	Escherichi	MNRTSPYY	RVNSLLLI	-	100.0	87	99.5	23	98.7	22	-	-	98.3	17	99.7	40	99.5	30	99.6	39	99.9	23	-	-	91.9	26	98.7	24	100.0	39	99.8	34	-	-	Self-Associating AutoTra	IPR030930:	Yes	AFTTNVIGV	1DAB;Cla	n.d.	n.d.	n.d.
YP_0022	Hypothetical	Escherichi	MNTIHLRCL	RVNSLLLI	-	99.9	127	99.5	22	98.7	23	-	-	98.2	22	99.7	37	99.6	34	99.6	38	99.9	23	-	-	94.3	27	98.8	21	-	-	99.8	32	-	-	Self-Associating AutoTra	IPR030930:	Yes	AFTPDVIGV	1DAB;Cla	n.d.	473-603	n.d.
YP_0030	Adhesin-like	Escherichi	MNRTSPYY	RVNSLLLI	-	100.0	84	99.6	23	98.8	22	-	-	98.4	22	99.7	38	99.6	33	99.6	35	100.0	23	-	-	95.8	28	98.8	26	100.0	37	99.9	32	-	-	Self-Associating AutoTra	IPR030930:	Yes	AFTPDVIGV	1DAB;Cla	n.d.	265-383	n.d.
NP_2892	Abc transpor	Escherichi	MNRTSPYY	RVNSLLLI	-	100.0	84	99.6	22	98.7	18	-	-	98.3	17	99.7	39	99.6	33	99.6	39	99.9	23	-	-	94.3	27	98.8	21	100.0	35	99.8	32	-	-	Self-Associating AutoTra	IPR030930:	Yes	AFTPDVIGV	1DAB;Cla	n.d.	n.d.	n.d.
NP_8851	Autotranspor	Bordetella	MLRSVSPLF	AASSVGE	-	100.0	63	98.8	22	97.9	20	-	-	97.2	16	99.3	40	98.8	21	98.8	25	100.0	21	-	-	95.0	24	98.9	21	100.0	41	99.8	24	-	-	Autotransporter of unknc	IPR005546:	Yes	EHWSGATI	1DAB;Cla	n.d.	n.d.	L1 (95-17
NP_8894	Autotranspor	Bordetella	MLRSVSPLF	AASSVGE	-	100.0	68	98.8	16	98.0	21	-	-	97.2	19	99.3	40	98.8	20	98.8	25	99.9	20	-	-	97.6	23	99.2	21	100.0	42	99.8	26	-	-	Autotransporter of unknc	IPR005546:	Yes	EHWSGATI	1DAB;Cla	n.d.	428-554	L1 (95-17
NP_8834	Pertactin pre	Bordetella	MNMSLSRIV	DNSNVGA	-	100.0	103	99.0	23	98.2	25	-	-	97.8	19	99.5	95	99.0	34	99.0	28	100.0	21	-	-	94.6	30	98.9	23	100.0	95	99.8	35	-	-	Autotransporter of unknc	IPR005546:	Yes	DWNNQSII	1DAB;Cla	n.d.	n.d.	L1 (112-1
NP_8879	Pertactin pre	Bordetella	MNMSLSRIV	DNSNVGA	-	100.0	93	99.2	23	98.3	28	-	-	97.6	16	99.6	96	99.2	34	99.2	25	100.0	21	-	-	88.8	26	98.8	24	100.0	96	99.8	35	-	-	Autotransporter of unknc	IPR005546:	Yes	DWNNQSII	1DAB;Cla	n.d.	262-400	L1 (165-2
NP_8798	Pertactin aut	Bordetella	MNMSLSRIV	DNSNVGA	-	100.0	89	99.1	24	98.2	29	-	-	97.6	16	99.6	100	99.1	30	99.1	32	100.0	21	-	-	88.1	28	98.8	23	100.0	100	99.8	36	-	-	Autotransporter of unknc	IPR005546:	Yes	DWNNQSIV	1DAB;Cla	n.d.	n.d.	L1 (162-2
NP_8813	Autotranspor	Bordetella	MNDRKSNS	TVTGSST	-	100.0	62	99.7	25	99.4	23	95.0	24	99.2	19	99.9	53	99.8	38	99.7	29	99.9	25	-	-	97.7	25	99.3	23	100.0	47	99.8	38	-	-	Autotransporter of unknc	IPR005546:	Yes	QAAPAAAE	1DAB;Cla	n.d.	n.d.	L1 (266-3
NP_8885	Autotranspor	Bordetella	MNDRKSNS	TVTGSSA	-	100.0	54	99.8	25	99.6	26	96.5	20	99.4	17	99.9	54	99.8	36	99.8	28	99.9	23	-	-	97.7	23	99.3	22	100.0	48	99.9	37	-	-	Autotransporter of unknc	IPR005546:	Yes	QAAPAAAA	1DAB;Cla	n.d.	262-400	L1 (272-3
NP_8848	Autotranspor	Bordetella	MNDRKSNS	TVTGSSA	-	100.0	63	99.3	23	98.8	25	34.1	22	98.5	17	99.7	49	99.5	36	99.5	28	99.9	24	-	-	97.5	23	99.2	22	100.0	48	99.8	36	-	-	Autotransporter of unknc	IPR005546:	Yes	QAAPAAAA	1DAB;Cla	n.d.	n.d.	L1 (264-3
YP_0030	Putative adhe	Escherichi	MGIKQHNGI	NSNLDTL	-	100.0	93	98.9	20	98.2	21	-	-	97.5	20	98.8	33	98.4	35	98.7	27	100.0	19	-	-	94.8	30	99.0	26	100.0	32	99.8	32	-	-	Autotransporter of unknc	IPR005546:	Yes	ADTGSQY	1DAB;Cla	n.d.	322-432	n.d.
YP_0014	Outer membr	Escherichi	MGIKQHNGI	NSNLDTL	-	100.0	91	97.8	18	95.3	21	-	-	91.2	18	97.5	33	96.1	23	97.0	26	100.0	19	-	-	95.4	29	99.0	25	100.0	32	99.8	33	-	-	Autotransporter of unknc	IPR005546:	Yes	ADTGSQY	1DAB;Cla	n.d.	428-554	L1 (160-2
YP_0017	Outer membr	Escherichi	MGIKQHNGI	NSNLDTL	-	100.0	95	98.7	18	97.7	21	-	-	97.0	20	98.5	33	98.0	33	98.3	25	100.0	21	-	-	94.8	31	98.9	25	100.0	32	99.8	36	-	-	Autotransporter of unknc	IPR005546:	Yes	ADTGSQY	1DAB;Cla	n.d.	428-554	n.d.
NP_4157	Putative adhe	Escherichi	MGIKQHNGI	NSNLDTL	-	100.0	91	97.8	19	92.8	21	-	-	87.6	19	97.5	33	95.5	32	96.9	25	100.0	19	-	-	94.8	30	99.0	26	100.0	32	99.8	32	-	-	Autotransporter of unknc	IPR005546:	Yes	ADTGSQY	1DAB;Cla	n.d.	n.d.	n.d.
YP_0029	Putative adhe	Escherichi	MGIKQHNGI	NSNLDTL	-	100.0	91	97.8	19	92.8	21	-	-	87.6	19	97.5	33	95.5	32	96.9	25	100.0	19	-	-	91.5	29	98.8	23	100.0	31	99.8	34	-	-	Autotransporter of unknc	IPR005546:	Yes	ADTGSQY	1DAB;Cla	n.d.	473-603	n.d.
YP_0024	Adhesin; aut	Escherichi	MGIKQHNGI	NSNLDTL	-	100.0	95	98.6	17	97.5	21	-	-	96.1	20	98.3	33	97.7	34	98.1	25	100.0	21	-	-	94.8	31	98.9	25	100.0	32	99.8	36	-	-	Autotransporter of unknc	IPR005546:	Yes	ADTGSQY	1DAB;Cla	n.d.	473-603	L1 (160-2
YP_0022	Putative auto	Escherichi	MGIKQHNGI	NSNLDTL	-	100.0	97	98.7	17	97.8	21	-	-	96.7	20	98.5	33	98.0	34	98.4	25	100.0	21	-	-	94.8	31	98.9	25	100.0	32	99.8	36	-	-	Autotransporter of unknc	IPR005546:	Yes	ADTGSQY	1DAB;Cla	n.d.	473-603	L1 (160-2
YP_0032	Adhesin	Escherichi	MGIKQHNGI	NSNLDTL	-	100.0	96	98.7	19	97.5	21	-	-	96.5	20	98.2	33	97.7	33	98.1	25	100.0	19	-	-	94.8	30	99.0	26	100.0	32	99.8	32	-	-	Autotransporter of unknc	IPR005546:	Yes	ADTGSQY	1DAB;Cla	n.d.	265-383	n.d.
NP_8809	Autotranspor	Bordetella	MAGQARGV	GVWTVTG	-	100.0	64	97.2	21	95.0	12	-	-	92.9	14	98.0	27	97.7	25	97.6	20	99.9	22	-	-	97.7	19	99.4	18	100.0	28	99.8	25	-	-	Autotransporter of unknc	IPR005546:	Yes	AQRIDGGA	1DAB;Cla	n.d.	n.d.	L1 (67-14
NP_8884	Autotranspor	Bordetella	MAGQARRV	DSRVAEV	-	100.0	68	97.2	15	95.2	13	-	-	89.9	13	97.9	26	97.6	25	97.5	18	99.9	26	-	-	96.1	22	98.6	17	100.0	27	99.8	30	-	-	Autotransporter of unknc	IPR005546:	Yes	AQRIDGGA	1DAB;Cla	n.d.	262-400	L1 (67-14
NP_8846	Autotranspor	Bordetella	MAGQARRV	DSRVAEV	-	100.0	68	97.2	15	95.2	13	-	-	89.9	13	97.9	26	97.6	25	97.5	18	99.9	26	-	-	96.1	22	98.6	17	100.0	27	99.8	30	-	-	Autotransporter of unknc	IPR005546:	Yes	AQRIDGGA	1DAB;Cla	n.d.	n.d.	n.d.
YP_2754	Outer membr	Pseudomo	MSSRKSNV	NAVRSLS	-	100.0	48	31.7	26	-	-	-	-	-	-	-	-	-	-	31.8	28	99.9	26	-	-	96.7	22	98.8	18	100.0	39	99.8	30	-	-	Autotransporter of unknc	IPR005546:	Yes	AASSDNN	1DAB;Cla	n.d.	180-297	L1 (194-2
NP_7918	Autotranspor	Pseudomo	MVSYRKSN	DNAVTSL	-	100.0	45	94.7	27	58.1	17	-	-	56.1	12	91.4	41	84.6	27	93.5	30	100.0	23	-	-	96.8	22	98.9	22	100.0	40	99.8	29	-	-	Autotransporter of unknc	IPR005546:	Yes	DIQRKVNA	1DAB;Cla	n.d.	n.d.	T3 (291-3
YP_2364	Outer membr	Pseudomo	MADSPLREI	VRSLSLQ	-	100.0	116	64.0	22	-	-	-	-	-	-	-	-	38.9	28	56.6	30	99.9	25	-	-	96.0	23	98.6	18	100.0	41	99.8	32	-	-	Autotransporter of unknc	IPR005546:	Yes	HSVVEGK	1DAB;Cla	n.d.	180-297	L1 (166-2
YP_0012	Outer membr	Pseudomo	MGIVKKQR	VGSLALE	-	100.0	46	50.9	16	-	-	-	-	-	-	76.4	42	71.7	30	81.6	30	99.9	24	-	-	95.5	25	98.7	19	100.0	46	99.9	29	-	-	Autotransporter of unknc	IPR005546:	Yes	LTTVDGDL	1DAB;Cla	n.d.	428-554	L1 (90-15
YP_0016	Outer membr	Pseudomo	MIKRRNFTL	VAQLAMA	-	100.0	51	34.7	16	-	-	-	-	-	-	-	-	-	-	-	-	99.9	22	-	-	96.0	26	98.8	22	100.0	40	99.9	30	-	-	Autotransporter of unknc	IPR005546:	Yes	LTLVDDEQ	1DAB;Cla	n.d.	428-554	L1 (142-2
YP_2351	Outer membr	Pseudomo	MNAPFVLRF	GQWVMV	-	100.0	47	95.6	14	75.5	16	-	-	82.8	18	96.1	33	96.2	30	96.5	33	99.9	18	-	-	98.2	28	99.4	20	100.0	39	99.9	29	-	-	Autotransporter of unknc	IPR005546:	Yes	QILVNAPQ	1DAB;Cla	n.d.	67-172	L1 (140-2
YP_0017	Outer membr	Pseudomo	MPSFSPVTL	LNDLSMD	-	100.0	45	30.5	21	-	-	-	-	-	-	35.5	38	-	-	27.7	28	99.9	23	-	-	96.8	26	98.9	22	100.0	39	99.8	31	-	-	Autotransporter of unknc	IPR005546:	Yes	QTIRDDTLI	1DAB;Cla	n.d.	428-554	L1 (89-15
NP_7440	Outer membr	Pseudomo	MPSRLFHAL	WTLVDNS	-	100.0	49	79.2	18	-	-	-	-	-	-	86.3	37	87.4	30	90.3	30	99.9	23	-	-	98.5	23	99.5	25	100.0	41	99.9	30	-	-	Autotransporter of unknc	IPR005546:	Yes	ETLVDADT	1DAB;Cla	n.d.	n.d.	L1 (177-2
YP_3486	Outer membr	Pseudomo	MQRKISNVF	IENLTLNG	-	100.0	51	20.6	18	-	-	-	-	-	-	66.8	37	23.5	29	57.6	24	100.0	25	-	-	95.8	26	98.8	24	100.0	38	99.8	30	-	-	Autotransporter of unknc	IPR005546:	Yes	GNLIGGTT	1DAB;Cla	n.d.	180-297	L1 (74-15
NP_7920	Autotranspor	Pseudomo	MMLRPPCA	WAMIGNG	-	100.0	51	95.4	16	81.2	17	-	-	82.3	22	96.2	33	96.3	33	96.7	32	99.9	19	-	-	98.2	33	99.5	22	100.0	37	99.9	34	-	-	Autotransporter of unknc	IPR005546:	Yes	ATVPVKDE	1DAB;Cla	n.d.	n.d.	L1 (103-1
YP_0028	Hypothetical	Pseudomo	MSTEKTFLL	LSMNGGN	-	100.0	44	40.4	21	-	-	-	-	-	-	-	-	-	-	45.3	26	100.0	26	-	-	95.7	22	98.8	21	100.0	36	99.8	30	-	-	Autotransporter of unknc	IPR005546:	Yes	ATQVKALN	1DAB;Cla	n.d.	473-603	L1 (89-15
YP_6097	Autotranspor	Pseudomo	MVEGLLGS	LTGNSDV	-	100.0	78	97.7	16	94.9	19	-	-	94.6	29	97.8	34	97.3	30	98.0	33	99.9	21	-	-	97.8	27	99.3	33	100.0	42	99.8	32	-	-	Autotransporter of unknc	IPR005546:	Yes	VVLNNQTL	1DAB;Cla	n.d.	180-297	L1 (117-1
YP_6472	Autotranspor	Yersinia pe	MNANADSC	AYATVDN	-	100.0	164	98.4	21	97.9	15	-	-	96.7	13	98.8	40	98.4	24	98.5	21	100.0	23	-	-	94.8	22	98.8	14	100.0	39	99.8	30	-	-	Autotransporter of unknc	IPR005546:	Yes	MVDQIFVT	1DAB;Cla	n.d.	180-297	n.d.
YP_0712	Pertactin fam	Yersinia ps	MNANADSC	AYATVDN	-	100.0	164	98.5	21	98.0	14	-	-	96.6	12	98.9	40	98.6	24	98.7	22	99.9	23	-	-	94.1	22	98.8	14	100.0	39	99.8	30	-	-	Autotransporter of unknc	IPR005546:	Yes	MVDQIFVT	1DAB;Cla	n.d.	91-203	n.d.
YP_6521	Putative auto	Yersinia pe	MNANADSC	AYATVDN	-	100.0	164	98.4	21	97.9	15	-	-	96.7	13	98.8	40	98.4	24	98.5	21	100.0	23	-	-	94.8	22	98.8	14	100.0	39	99.8	30	-	-	Autotransporter of unknc	IPR005546:	Yes	MVDQIFVT	1DAB;Cla	n.d.	180-297	n.d.
YP_0035	Putative auto	Yersinia pe	MNTNSKKT	AYATVDN	-	100.0	62	97.0	21	96.0	14	-	-	89.9	13	97.9	39	97.6	24	97.5	22	100.0	22	-	-	92.8	23	98.8	14	100.0	38	99.8	32	-	-	Autotransporter of unknc	IPR005546:	Yes	DSCIDYTV	1DAB;Cla	n.d.	200-322	n.d.
YP_0016	Putative auto	Yersinia pe	MNTNSKKT	AYATVDN	-	100.0	62	97.0	21	96.0	14	-	-	89.9	13	97.9	39	97.6	24	97.5	22	100.0	22	-	-	92.8	23	98.8	14	100.0	38	99.8	32	-	-	Autotransporter of unknc	IPR005546:	Yes	DSCIDYTV	1DAB;Cla	n.d.	428-554	n.d.
YP_0018	Outer membr	Yersinia ps	MNTNSKKT	AYATVDN	-	100.0	58	97.7	22	96.6	13	-	-	93.0	13	98.2	40	97.8	31	97.9	23	100.0	22	-	-	92.5	23	98.7	14	100.0	38	99.8	31	-	-	Autotransporter of unknc	IPR005546:	Yes	ANADSCTD	1DAB;Cla	n.d.	428-554	n.d.
YP_0017	Outer membr	Yersinia ps	MNTNSKKT	AYATVDN	-	100.0	58	98.3	20	97.6	13	-	-	95.7	13	98.6	39	98.4	31	98.4	23	100.0	22	-	-	95.4	23	98.9	13	100.0	39	99.8	30	-	-	Autotransporter of unknc	IPR005546:	Yes	ANADSCTD	1DAB;Cla	n.d.	428-554	n.d.
YP_0014	Autotranspor	Yersinia ps	MDKTLLAG	TRINNVTL	-	100.0	54	97.2	15	96.5	14	-	-	91.1	11	98.2	35	97.9	32	97.9	34	99.9	18	-	-	93.0	27	98.7	18	100.0	34	99.8	34	-	-	Autotransporter of unknc	IPR005546:	Yes	FTPNVVGI	1DAB;Cla	n.d.	428-554	T4 (70-12
YP_0018	Outer membr	Yersinia ps	MDKTLLAG	TRINNVTL	-	100.0	54	97.6	16	97.2	16	-	-	95.2	12	98.7	35	98.4	27	98.3	34	99.9	18	-	-	93.0	27	98.7	18	100.0	34	99.8	34	-	-	Autotransporter of unknc	IPR005546:	Yes	FTPNVVGI	1DAB;Cla	n.d.	428-554	n.d.
YP_0017	Outer membr	Yersinia ps	MDKTLLAG	TRINNVTL	-	100.0	54	97.2	15	96.5	14	-	-	91.1	11	98.2	35	97.9	32	97.9	34	99.9	19	-	-	95.1	25	98.7	17	100.0	34	99.8	33	-	-	Autotransporter of unknc	IPR005546:	Yes	FTPNVVGI	1DAB;Cla	n.d.	428-554	n.d.
YP_0695	Pertactin fam	Yersinia ps	MDKTLLAG	TRINNVTL	-	100.0	54	97.6	16	97.2	16	-	-	95.2	12	98.7	35	98.4	27	98.3	34	99.9	20	-	-	94.6	25	98.8	18	100.0	34	99.8	34	-	-	Autotransporter of unknc	IPR005546:	Yes	FTPNVVGI	1DAB;Cla	n.d.	91-203	n.d.
YP_6469	Putative auto	Yersinia pe	MDKTLLAG	TRINNVTL	-	100.0	54	97.5	17	96.7	16	-	-	88.5	12	98.6	35	98.2	32	98.2	34	99.9	18	-	-	93.0	27	98.7	18	100.0	34	99.8	34	-	-	Autotransporter of unknc	IPR005546:	Yes	FTPNVVGI	1DAB;Cla	n.d.	180-297	n.d.
YP_0011	Putative auto	Yersinia pe	MDKTLLAG	TRINNVTL	-	100.0	54	97.5	17	96.7	16	-	-	88.5	12	98.6	35	98.2	32	98.2	34	99.9	18	-	-	93.0	27	98.7	18	100.0	34	99.8	34	-	-	Autotransporter of unknc	IPR005546:	Yes	FTPNVVGI	1DAB;Cla	n.d.	428-554	n.d.
NP_6684	Autotranspor	Yersinia pe	MDKTLLAG	TRINNVTL	-	100.0	54	97.5	17	96.7	16	-	-	88.5	12	98.6	35	98.2	32	98.2	34	99.9	20	-	-	91.0	27	98.7	18	100.0	34	99.8	34	-	-	Autotransporter of unknc	IPR005546:	Yes	FTPNVVGI	1DAB;Cla	n.d.	n.d.	n.d.
YP_6524	Putative auto	Yersinia pe	MDKTLLAG	TRINNVTL	-	100.0	54	97.5	17	96.7	16	-	-	88.5	12	98.6	35	98.2	32	98.2	34	99.9	18	-	-	93.0	27	98.7	18	100.0	34	99.8	34	-	-	Autotransporter of unknc	IPR005546:	Yes	FTPNVVGI	1DAB;Cla	n.d.	180-297	n.d.
YP_0035	Putative auto	Yersinia pe	MDKTLLAG	TRINNVTL	-	100.0	54	97.5	17	96.7	16	-	-	88.5	12	98.6	35	98.2	32	98.2	34	99.9	18	-	-	93.0	27	98.7	18	100.0	34	99.8	34	-	-	Autotransporter of unknc	IPR005546:	Yes	FTPNVVGI	1DAB;Cla	n.d.	200-322	n.d.
NP_9925	Autotranspor	Yersinia pe	MDKTLLAG	TRINNVTL	-	100.0	54	97.5	17	96.7	16	-	-	88.5	12	98.6	35	98.2	32	98.2	34	99.9	20	-	-	91.0	27	98.7	18	100.0	34	99.8	34	-	-	Autotransporter of unknc	IPR005546:	Yes	FTPNVVGI	1DAB;Cla	n.d.	428-554	n.d.
YP_0023	Autotranspor	Yersinia pe	MDKTLLAG	TRINNVTL	-	100.0	54	97.5	17	96.7	16	-	-	88.5	12	98.6	35	98.2	32	98.2	34	99.9	20	-	-	91.0	27	98.7	18	100.0	34	99.8	34	-	-	Autotransporter of unknc	IPR005546:	Yes	FTPNVVGI	1DAB;Cla	n.d.	473-603	n.d.
YP_0016	Putative auto	Yersinia pe	MDKTLLAG	TRINNVTL	-	100.0	55	97.3	18	96.6	16	-	-	87.5	12	98.5	35	98.0	32	98.1	34	99.9	19	-	-	94.0	25	98.7	17	100.0	34	99.8	34	-	-	Autotransporter of unknc	IPR005546:	Yes	FTPNVVGI	1DAB;Cla	n.d.	428-554	n.d.
NP_7949	Autotranspor	Pseudomo	MWTTGATW	GGINVAQ	-	97.5	97	20.1	23													97.5	21	-	-	3.1	18	87.6	14	97.2	17	86.7	10	-	-	Phosphata Phosphatase	IPR005546:	Yes	AGAFTTITC	1R6V;Clas	n.d.	n.d.	L1 (476-5
YP_2726	Autotranspor	Pseudomo	MRCNKPVA	GGINVAQ	-	97.3	110	23.0	24	-	-	-	-	-	-	-	-	-	-	-	-	97.3	24	-	-	5.6	16	78.5	14	96.6	17	62.1	9	-	-	Phosphata Phosphatase	IPR005546:	Yes	ASRANDDF	2AKC;Cla	n.d.	180-297	L1 (522-6
YP_0032	Outer membr	Pectobacte	MAFPTKALL	GNGSSFT	-	97.7	91	96.1	20.0	-	-	-	-	29.6	15	69.0	18	53.6	15	-	-	97.7	16	-	-	-	-	52.4	16	96.9	19	7.7	17	-	-	Phosphata Phosphatase	IPR005546:	Yes	AEAEPASA	1R6V;Clas	n.d.	265-383	n.d.
YP_0034	Outer membr	Azospirillu	MKGRSLTLL	GGLTVGN	-	97.4	105	30.9	19	-	-	-	-	-	-	-	-	-	-	-	-	97.4	20	-	-	-	-	54.3	16	95.7	17	25.6	10	-	-	Phosphata Phosphatase	IPR005546:	Yes	QQIPTSPS	1D2T;Clas	n.d.	265-383	T5 (565-6
YP_0015	Outer membr	Azorhizobi	MEGEVRRL	TLMPGTL	-	96.0	150	29.4	17	-	-	-	-	-	-	39.4	18	70.2	15	-	-	96.0	16	-	-	-	-	67.2	15	94.6	10	51.0	14	-	-	Phosphata Phosphatase	IPR005546:	Yes	QQVPTSPF	1D2T;Clas	n.d.	428-554	T5 (589-6
YP_0012	Serine protea	Bradyrhizo	MSSSAYLKF	PISVGAG	-	97.3	149	28.8	18	-	-	-	-	-	-	29.9	18	27.5	13	-	-	97.3	23	-	-	3.0	15	84.5	14	96.0	12	87.1	12	-	-	Phosphata Phosphatase	IPR005546:	Yes	GDIPTTPP	2AKC;Cla	n.d.	428-554	n.d.
YP_0012	Outer membr	Bradyrhizo	MSSSAYLK	KLNVLGS	-	92.3	147	41.2	15	-	-	-	-	-	-	33.9	19	38.7	16	-	-	92.3	18	-	-	19.3	15	90.1	16	87.3	16	55.2	14	-	-	Phosphata Phosphatase	IPR005546:	Yes	GDIPTTPP	2AKC;Cla	n.d.	428-554	L3 (556-6
YP_0036	Acid phospha	Moraxella	MPKYKLLSA	VTGQFQS	-	96.5	77	96.5	16.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Phosphata Phosphatase	IPR005546:	Yes	HDEAFDKL	1R6V;Clas	n.d.	200-322	n.d.
YP_0021	Autotranspor	Proteus mi	MAYFMKNIV	GKINKLAV	-	97.9	77	97.9	19.0	88.6	12	-	-	91.6	12	93.1	21	93.7	14	79.2	15	95.8	16	-	-	-	-	33.3	14	80.0	17	-	-	-	-	Phosphata Phosphatase	IPR001011:	Yes	ETNISQTK	1D2T;Clas	n.d.	132-263	L1 (461-5
YP_0639	Serine protea	Desulfotale	MRKTKLYLA	NYGTLAG	-	97.3	63	97.3	13.0	90.1	13	-	-	78.9	21	96.5	19	96.5	22	92.8	17	96.2	15	-	-	-	-	36.8	28	75.6	19	30.6	18	-	-	Autotransporter of unknc	IPR005546:	Yes	YETITVGNF	1SAT;Clas	IPR011050	91-203	n.d.
YP_0639	Serine protea	Desulfotale	MKASTIGMS	GSIIGDVD	-	97.2	72	97.2	15.0	73.3	13	-	-	81.3	18	93.1	20	95.2	21	82.7	18	96.2	22	-	-	-	-	54.0	23	71.0	21	25.4	18	-	-	Autotransporter of unknc	IPR005546:	Yes	YETITVGNF	1LML;Clas	IPR011050	91-203	n.d.
YP_0041	Outer membr	Desulfobul	MNLTLYASL	GETTVGA	-	90.1	48	86.7	21	-	-	-	-	48.8	25	62.1	14	67.3	16	-	-	90.1	21	-	-	-	-	59.2	28	46.2	16	59.4	14	-	-	Protease / Cystein peptic	IPR005546:	Yes	MDLPSSFD	1JQP;Clas	n.d.	91-203	n.d.
YP_0041	Outer membr	Desulfobul	MLPAVQAA	VGEVVSS	-	96.5	131	41.9	16	-	-	-	-	-	-	62.3	19	44.3	17	-	-	96.5	21	-	-	-	-	85.2	15	91.4	20	89.7	16	-	-	Protease / Cystein peptic	IPR005546:	Yes	STWNRAH	2O6X;Clas	n.d.	91-203	n.d.
YP_1632	Outer membr	Zymomona	MTKRKLSKF	DYAIITGLI	-	98.0	104	98.0	13	96.0	20	-	-	91.1	14	95.2	19	93.9	21	85.8	19	97.9	14	-	-	2.7	16	61.1	13	96.7	13	59.5	19	-	-	Protease / Peptidase S1	IPR005546:	Yes	IVINDNETP	4O03;Clas	n.d.	91-203	n.d.
YP_0032	Outer membr	Zymomona	MTKRKLSKF	DYAIITGLI	-	98.4	104	98.4	14	96.6	19	-	-	94.6	12	97.2	14	96.1	24	94.3	17	97.9	17	-	-	3.0	16	64.2	14	96.9	13	58.5	17	-	-	Protease / Peptidase S1	IPR005546:	Yes	IVINDNETP	4O03;Clas	n.d.	265-383	n.d.
YP_7580	Outer membr	Maricaulis	MPHTPTPLT	GSGTINTF	-	97.4	98	97.4	12.0	94.2	16	-	-	89.9	13	93.4	15	95.3	15	87.0	15	97.0	12	-	-	3.4	14	80.3	15	94.1	14	47.5	12	-	-	Protease / Peptidase S1	IPR005546:	Yes	QEEAPTVIF	4O03;Clas	n.d.	180-297	n.d.
YP_0033	Outer membr	Sebaldella	MKKKLILIAL	INNGGIISI	-	93.9	106	24.2	11	-	-	-	-	-	-	-	-	-	-	-	-	93.9	19	-	-	-	-	-	-	38.2	18	-	-	-	-	Protease / Peptidase S8	IPR005546:	Yes	GQSNQNPS	1R6V;Clas	n.d.	265-383	n.d.
YP_0033	Outer membr	Sebaldella	MNKKIMFIIT	NNNVVNN	-	95.4	101	73.1	12	-	-	-	-	-	-	-	-	-	-	-	-	95.4	20	-	-	2.8	17	-	-	15.5	17	2.6	17	-	-	Protease / Peptidase S8	IPR005546:	Yes	GGEGNSG	1R6V;Clas	n.d.	265-383	n.d.
YP_7861	Autotranspor	Bordetella	VQREEAPIE	SVTNNGA	-	96.4	270	69.5	15	-	-	-	-	-	-	-	-	-	-	-	-	96.4	20	-	-	3.4	14	-	-	86.4	20	7.8	20	-	-	Protease / Peptidase S8	IPR005546:	Yes	LSRAAMIA	1R6V;Clas	n.d.	180-297	n.d.
YP_0015	Autotranspor	Neisseria r	MRTTPTFPT	LNSDGIVY	-	94.1	141	20.4	23	-	-	-	-	-	-	-	-	-	-	-	-	94.1	27	-	-	-	-	-	-	89.5	17	-	-	-	-	Protease / Peptidase S8	IPR005546:	Yes	GGGGGGT	1R6V;Clas	n.d.	428-554	n.d.
YP_0019	Serine protea	Xanthomo	MKKKDVAR	GGQVDNA	-	95.7	53	53.9	15	-	-	-	-	-	-	-	-	-	-	-	-	95.7	22	-	-	-	-	-	-	87.4	21	13.5	10	-	-	Protease / Peptidase S8	IPR005546:	Yes	GGGGGIAF	1R6V;Clas	n.d.	132-263	TS (497-5
YP_2005	Serine protea	Xanthomo	MKKKDVAR	GGQVDNA	-	95.2	53	53.9	15	-	-	-	-	-	-	-	-	-	-	-	-	95.2	22	-	-	2.3	16	7.3	19	87.5	21	14.9	10	8.2	24.0	Protease / Peptidase S8	IPR005546:	Yes	GGGGGIAF	1R6V;Clas	n.d.	91-203	TS (497-5
YP_4508	Serine protea	Xanthomo	MKKKDVAR	GGQVDNA	-	95.2	53	53.9	15	-	-	-	-	-	-	-	-	-	-	-	-	95.2	22	-	-	2.3	16	7.3	19	87.5	21	14.9	10	8.2	24.0	Protease / Peptidase S8	IPR005546:	Yes	GGGGGIAF	1R6V;Clas	n.d.	180-297	TS (497-5
YP_3631	Autotranspor	Xanthomo	MKKQDVAR	GGRVDNA	-	95.2	61	27.0	15	-	-	-	-	-	-	-	-	-	-	-	-	95.2	22	-	-	3.6	14	45.7	20	85.4	24	22.6	10	-	-	Protease / Peptidase S8	IPR005546:	Yes	GGGGGIAF	1R6V;Clas	n.d.	180-297	TS (499-5
YP_2440	Serine protea	Xanthomo	MGPIHGLVF	AGTLNNA	-	96.8	99	24.4	18	-	-	-	-	-	-	-	-	-	-	-	-	96.8	26	-	-	-	-	-	-	90.9	22	61.7	16	-	-	Protease / Peptidase S8	IPR005546:	Yes	GGGGGGG	1R6V;Clas	n.d.	180-297	TS (480-5
NP_6366	Serine protea	Xanthomo	MGPIHGLVF	AGTLNNA	-	97.0	99	24.4	18	-	-	-	-	-	-	-	-	-	-	-	-	97.0	26	-	-	35.3	18	36.9	20	93.4	22	66.1	16	-	-	Protease / Peptidase S8	IPR005546:	Yes	GGGGGGG	1R6V;Clas	n.d.	n.d.	TS (480-5
YP_0033	Autotranspor	Xanthomo	MRTLMARS	INVLNAGT	-	95.4	103	52.6	13	-	-	-	-	-	-	-	-	-	-	-	-	95.4	12	-	-	6.6	14	41.1	12	50.5	20	9.5	17	-	-	Protease / Peptidase S8	IPR005546:	Yes	LTSCGGGH	1R6V;Clas	n.d.	265-383	L1 (427-5
YP_0017	Serine protea	Xylella fast	MKNMKTTF	GGAVSNA	-	95.8	62	76.9	18	-	-	-	-	-	-	-	-	-	-	-	-	95.8	16	-	-	6.0	11	-	-	58.2	17	4.1	16	-	-	Protease / Peptidase S8	IPR005546:	Yes	GSGLVRPF	1R6V;Clas	n.d.	428-554	n.d.
YP_0041	Autotranspor	Pseudoxar	MGDHLSVIF	LVNAGVV	-	96.5	167	27.4	16	-	-	-	-	-	-	-	-	-	-	-	-	96.5	22	-	-	8.0	11	42.3	16	94.4	15	45.2	13	-	-	Protease / Peptidase S8	IPR005546:	Yes	QWPGGVA	1R6V;Clas	n.d.	91-203	TS (356-3
YP_0019	Autotranspor	Stenotroph	MERTMMVR	IGNVVNG	-	96.8	59	31.2	17	-	-	-	-	-	-	-	-	-	-	-	-	96.8	20	-	-	55.5	16	89.5	16	86.9	20	9.8	14	-	-	Protease / Peptidase S8	IPR005546:	Yes	GGGGGGS	1R6V;Clas	n.d.	132-263	TS (504-5
YP_0041	Outer membr	Pseudoxar	MNKKFPIRG	PTFVGAN	-	97.8	100	78.9	19	30.7	13	-	-	48.6	15	-	-	26.4	16	-	-	97.8	22	-	-	28.9	21	72.4	17	96.4	16	34.8	19	-	-	Protease / Peptidase S8	IPR005546:	Yes	ATFDSLEQ	1R6V;Clas	n.d.	91-203	n.d.
YP_2603	Autotranspor	Pseudomo	MVVVKQFQ	GGLTANA	-	95.8	87	67.0	20	-	-	-	-	49.0	16	59.5	17	31.5	18	-	-	95.8	18	-	-	-	-	63.6	16	83.2	19	8.4	17	-	-	Protease / Peptidase S8	IPR005546:	Yes	APYVETGK	1R6V;Clas	n.d.	180-297	n.d.

YP_3484Outer membrPseudomoMKNNNTPA/AALTANR	-	94.2	63	-	-	-	-	-	-	22.0	11	-	-	24.5	14	-	-	94.2	18	-	-	-	-	81.1	12	75.7	17	12.1	19	-	-	Protease / Peptidase S8	IPR005546:	Yes	APYVESGF	1R6V;Cla	n.d.	180-297	n.d.
YP_0014Outer membrSerratia prMKQATGKNGSLTARS	-	96.6	101	20.4	28	-	-	-	-	-	-	-	-	-	-	-	-	96.6	25	-	-	-	-	52.8	13	45.3	26	-	-	-	-	Protease / Peptidase S8	IPR005546:	Yes	APYTEGGK	1R6V;Cla	n.d.	428-554	n.d.
YP_0037Extracellular XenorhabdMKKRYRIAFGELNVQR	-	96.8	111	43.6	14	-	-	-	-	-	-	-	-	-	-	-	-	96.8	23	-	-	-	-	67.5	9	77.7	14	-	-	-	-	Protease / Peptidase S8	IPR005546:	Yes	ENYVEKGF	1R6V;Cla	n.d.	200-322	n.d.
YP_0014Outer membrSerratia prMLFCLAAIGRSGGTVA	-	95.4	71	25.4	24	-	-	-	-	-	-	-	-	-	-	-	-	95.4	23	-	-	-	-	60.9	13	71.7	20	4.9	16	-	-	Protease / Peptidase S8	IPR005546:	Yes	TPYVESGK	1R6V;Cla	n.d.	428-554	n.d.
YP_0034Extracellular XenorhabdMKKNNQLRGSLTAYN	-	96.7	76	31.0	14	-	-	-	-	-	-	-	-	-	-	-	-	96.7	19	-	-	-	-	83.9	10	76.5	16	-	-	-	-	Protease / Peptidase S8	IPR005546:	Yes	ATEYHEQC	1R6V;Cla	n.d.	265-383	n.d.
YP_7896Serine proteaPseudomoMTDDHSFRIGGLLARS	-	95.4	83	24.9	16	-	-	-	-	-	-	-	-	35.3	21	-	-	95.4	24	-	-	-	-	60.6	13	80.4	17	6.9	17	-	-	Protease / Peptidase S8	IPR005546:	Yes	AYVEAGRF	1R6V;Cla	n.d.	474-584	n.d.
YP_0013Putative serinPseudomoMTDNRPRL'GSLSVEG	-	94.6	80	31.7	21	-	-	-	-	-	-	-	-	-	-	-	-	94.6	22	-	-	2.1	16	72.6	15	60.4	19	10.7	14	-	-	Protease / Peptidase S8	IPR005546:	Yes	AYVETGRP	1R6V;Cla	n.d.	428-554	n.d.
YP_0010Serine proteaCampylobaMKKSIFLSFIFSGNGEV	-	90.3	135	33.3	16	-	-	-	-	-	-	-	-	-	-	-	-	90.3	18	-	-	-	-	4.0	21	11.0	14	-	-	-	-	Protease / Peptidase S8	IPR005546:	Yes	DQELIGSKI	1R6V;Cla	n.d.	428-554	n.d.
YP_0040Serine proteaCampylobaMKKSIFLSFIGDVKILE	-	93.5	132	53.9	19	-	-	-	-	-	-	-	-	24.1	13	-	-	93.5	17	-	-	6.8	17	12.0	17	-	-	-	-	-	-	Protease / Peptidase S8	IPR005546:	Yes	DQELISGKI	1DAB;Cla	n.d.	91-203	n.d.
YP_0023Outer membrCampylobaMKKFFCLTLVGQNLNN	-	92.7	71	92.7	13	27.0	10	-	-	47.8	11	-	-	42.5	16	-	-	61.1	12	-	-	-	-	-	-	2.9	9	-	-	-	-	Protease / Peptidase S8	IPR005546:	Yes	LSEFELHHI	1R6V;Cla	n.d.	473-603	n.d.
YP_0010Serine proteaCampylobaMKKFFCLTLVGQNLNN	-	92.7	71	92.7	13	27.0	10	-	-	47.8	11	-	-	42.5	16	-	-	85.2	17	-	-	-	-	2.3	23	-	-	-	-	-	-	Protease / Peptidase S8	IPR005546:	Yes	LSEFELHHI	1R6V;Cla	n.d.	428-554	n.d.
YP_8929Serine proteaCampylobaMKKSDKMVSNVFVEQ	-	97.6	35	96.1	13	-	-	-	-	20.8	16	-	-	-	-	-	-	97.6	18	-	-	-	-	-	-	93.0	13	-	-	63.6	17.0	Protease / Peptidase S8	IPR005546:	Yes	LEPNFEKD	1R6V;Cla	n.d.	345-479	n.d.
YP_0012Hypothetical BradyrhizoMMLRSRTCFVITNYQN	-	95.9	52	95.9	9.0	-	-	-	-	-	-	27.6	13	-	-	40.3	11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Autotransporter of unknc	IPR005546:	Yes	APVRAQST	n.d.	n.d.	428-554	n.d.
NP_1026Hypothetical MesorhizoMTLTLFVGSDTVYLDL/	-	98.2	192	40.5	18	-	-	-	-	-	-	-	-	-	-	-	-	98.2	20	-	-	-	-	61.0	12	95.2	12	57.1	17	-	-	Autotransporter of unknc	IPR005546:	Yes	AGGPSGN	n.d.	n.d.	n.d.	L3 (2687-
YP_9114Outer membrChlorobiumMYGKLAHYILTNHGYV	-	97.4	150	35.6	19	-	-	-	-	-	-	-	-	-	-	-	-	97.4	17	-	-	-	-	57.8	13	84.9	17	-	-	-	-	Autotransporter of unknc	IPR005546:	Yes	ADPPPPPP	n.d.	n.d.	345-479	L3 (713-79
YP_0016Outer membrCaulobacteMGKLTSAA1IVTPGGN	-	97.2	223	69.6	13	-	-	-	-	-	-	-	-	-	-	-	-	97.2	15	-	-	-	-	41.2	11	87.0	16	9.1	19	-	-	Autotransporter of unknc	IPR005546:	Yes	GDYTQHAI	n.d.	n.d.	428-554	n.d.
YP_0041Outer membrMesorhizoMALFTERVAIDLGGGI	-	97.0	162	54.8	16	-	-	-	-	-	-	-	-	23.6	19	27.4	14	97.0	17	-	-	-	-	13.9	14	71.3	9	11.6	15	-	-	Autotransporter of unknc	IPR005546:	Yes	QQLSINTD	2ZJ6;Clas	n.d.	91-203	n.d.
YP_0015Outer membrAzorhizobiMSNVISRRAGGLTVAS	-	97.5	117	39.9	20	-	-	-	-	-	-	-	-	-	-	-	-	97.5	22	-	-	-	-	67.3	15	95.6	21	28.4	15	-	-	Autotransporter of unknc	IPR005546:	Yes	GCTTIGNQ	n.d.	n.d.	428-554	L1 (408-4
YP_9168Outer membrParacoccuMVTCNTIGNFHGLTVA	-	97.1	115	68.2	4	-	-	-	-	-	-	-	-	-	-	-	-	97.1	13	-	-	-	-	17.2	15	92.2	16	-	-	-	-	Autotransporter of unknc	IPR005546:	Yes	DPRSVPLD	n.d.	n.d.	345-479	L3 (479-5
YP_0017Outer membrMethylobaMGRGGRDFTQGTSFV	-	100.0	86	96.3	34	-	-	-	-	-	-	51.2	23	41.0	24	55.3	17	100.0	34	-	-	77.3	22	98.5	16	99.9	23	99.7	24	-	-	Autotransporter of unknc	IPR005546:	Yes	QSVVPGFD	n.d.	n.d.	428-554	T5 (579-6
YP_0024Outer membrMethylobaMLGCESGRTQPASVIN	-	100.0	99	98.5	30	92.3	22	-	-	91.9	13	97.2	24	96.8	25	96.4	18	100.0	30	-	-	69.0	22	98.6	13	99.9	25	99.7	24	-	-	Autotransporter of unknc	IPR005546:	Yes	VVPGFDAN	n.d.	n.d.	473-603	L3 (530-6
YP_0030Putative autoNeisseria mMAVDYVYDIRWDSDV	-	100.0	107	100.0	25	99.9	19	99.3	23	99.9	15	100.0	21	100.0	21	99.9	18	95.4	27	-	-	-	-	7.2	15	30.7	19	6.8	21	-	-	Autotransporter of unknc	IPR005546:	Yes	EEEKTLQQ	1DAB;Cla	n.d.	265-383	n.d.
YP_0029Type v secreCandidatuMTGHHLDINNLKNAGC	-	100.0	199	98.3	30	96.2	17	-	-	92.3	21	97.3	28	95.0	22	96.4	19	100.0	30	-	-	95.8	30	99.2	22	100.0	29	99.7	22	-	-	Autotransporter of unknc	IPR005546:	Yes	GGSLHAAC	1DAB;Cla	n.d.	473-603	n.d.
YP_0015Hypothetical SalmonellaMKIAPYLITLNYQIAGLE	-	100.0	30	79.4	12	-	-	-	-	-	-	95.0	39	87.1	28	91.6	26	99.9	17	-	-	97.1	20	99.0	21	100.0	37	99.9	25	-	-	Autotransporter of unknc	IPR005546:	Yes	DYHTQINA	1DAB;Cla	n.d.	428-554	T4 (133-2
YP_0010AutotransporYersinia eMHNSQHASNAEISALT	-	94.1	59	94.1	17	78.8	15	-	-	79.6	12	86.4	18	50.4	19	69.9	23	12.1	20	-	-	-	-	-	-	-	-	-	-	-	-	Autotransporter of unknc	IPR005546:	Yes	QTREFTYS	n.d.	n.d.	428-554	L1 (111-1
YP_0035AutotransporHelicobactMDRILAKCTGNPVINNI	-	99.7	117	99.6	14	99.6	22	98.6	19	99.5	22	99.7	19	99.6	19	99.6	34	97.1	19	-	-	-	-	42.3	23	96.7	24	49.1	20	-	-	Autotransporter of unknc	IPR005546:	Yes	ITITDANQII	1DAB;Cla	IPR011050;IPF	200-322	n.d.
YP_0035AutotransporHelicobactMNKKFHQPRIIIEGANC	-	96.7	191	96.7	13.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Autotransporter of unknc	IPR005546:	Yes	QASTTTSN	1DAB;Cla	n.d.	200-322	n.d.
YP_0041Hypothetical TaylorellaMNKIFKVVYYVVGGNF	-	100.0	64	99.5	27	97.9	21	-	-	97.9	22	98.4	22	98.3	21	98.3	20	100.0	27	-	-	5.0	24	98.7	24	99.9	18	99.2	20	-	-	Autotransporter of unknc	IPR005546:	Yes	QTTTVDLT	3C7H;Cla	n.d.	91-203	L1 (626-70
YP_9127Outer membrChlorobiumMKKRRLSL/GTCTQSG	-	99.8	41	99.8	23	99.3	15	87.2	16	99.2	16	99.7	20	99.6	21	99.6	16	98.1	27	-	-	-	-	19.6	14	95.9	23	33.4	24	-	-	Autotransporter of unknc	IPR005546:	Yes	LSVPVPTY	1DAB;Cla	n.d.	345-479	L1 (272-3
YP_0036Outer membrStarkeya nMAVAGGPGVSGIFTSV	-	97.6	76	73.7	20	-	-	-	-	-	-	49.2	19	-	-	-	-	97.6	32	-	-	-	-	-	-	96.9	23	9.7	22	-	-	Autotransporter of unknc	IPR005546:	Yes	GTTYSANT	1DAB;Cla	n.d.	200-322	L1 (745-8
YP_0041Hypothetical TaylorellaMNIKFKVVFLKGSLNN	-	99.8	67	99.8	20	98.8	17	-	-	98.8	19	99.4	18	99.2	17	99.3	22	99.1	23	-	-	-	-	94.9	16	98.2	15	53.7	17	-	-	Protease / Peptidase M1	IPR005546:	Yes	GYLTPDRP	n.d.	n.d.	91-203	n.d.
YP_0018Outer membrElusimicroMKNIFKITAINSSSDIFI	-	100.0	41	99.0	21	98.4	19	20.4	18	97.8	20	99.0	22	99.1	23	98.8	31	100.0	22	-	-	97.6	18	99.4	20	100.0	20	99.8	24	-	-	Autotransporter of unknc	IPR005546:	Yes	AQLVLTSAI	1DAB;Cla	n.d.	132-263	L1 (230-3
YP_0037Outer membrHyphomicrMGRRDLSTINTNGTIDL	-	99.9	87	99.5	23	98.7	17	-	-	98.4	16	99.2	26	99.1	19	99.1	17	99.9	30	-	-	72.7	19	98.5	19	99.9	27	99.6	26	-	-	Autotransporter of unknc	IPR005546:	Yes	AGCTSSVA	1DAB;Cla	n.d.	200-322	L1 (267-3
YP_0039Outer membrEnterobactMNKIFRVIWDVNLAGT	-	100.0	68	99.6	31	98.9	21	-	-	98.5	21	99.3	28	99.1	21	99.3	25	100.0	31	-	-	93.8	23	99.1	22	100.0	29	99.7	21	-	-	Autotransporter of unknc	IPR005546:	Yes	ADYTSTVT	1DAB;Cla	n.d.	91-203	L1 (110-1
NP_3851Hypothetical SinorhizobMIPAGPVLAYSISGNL	-	100.0	199	97.0	25	-	-	-	-	-	-	61.2	27	-	-	28.6	19	100.0	27	-	-	62.3	17	98.8	17	100.0	28	99.7	24	-	-	Autotransporter of unknc	IPR005546:	Yes	INECGPAIC	2AGM;Cla	n.d.	n.d.	L1 (293-3
YP_0042Outer membrRahnella sMKWKPRLLESDTFVA	-	100.0	117	98.9	26	97.4	13	-	-	96.4	16	98.0	20	98.3	24	97.5	18	100.0	29	-	-	3.1	22	98.8	19	99.9	19	99.4	23	-	-	Protease / Peptidase S6	IPR005546:	Yes	GIMREDVA	1DAB;Cla	n.d.	91-203	T6 (586-6
YP_0036Outer membrBurkholderMAARSIHRATVTNAGT	-	100.0	103	99.7	33	99.0	17	-	-	98.4	16	99.4	26	99.0	23	99.0	16	100.0	35	-	-	93.6	19	98.9	18	100.0	24	99.6	25	-	-	Autotransporter of unknc	IPR005546:	Yes	QQTVTERF	1DAB;Cla	n.d.	200-322	L3 (187-2
YP_2945Outer membrRalstonia MQRCSRFRQILGVRTF	-	100.0	188	98.2	37	91.4	19	-	-	86.3	17	94.5	29	93.3	19	94.3	20	100.0	33	-	-	-	-	98.5	17	99.9	24	99.4	21	-	-	Autotransporter of unknc	IPR005546:	Yes	GSGTLDGV	1DAB;Cla	n.d.	180-297	L3 (278-3
YP_0020Outer membrProsthecoMKKRLSLIMVGNLNNN	-	100.0	37	99.7	20	98.8	15	-	-	98.6	12	99.1	20	99.1	17	99.1	7	100.0	27	-	-	-	-	98.2	13	99.9	22	99.3	19	-	-	Autotransporter of unknc	IPR005546:	Yes	ADWTVSDF	1DAB;Cla	n.d.	132-263	n.d.
NP_9005Hypothetical ChromobaMKTYQQKQGGQTAG/	-	100.0	57	94.4	34	-	-	-	-	-	-	56.5	22	-	-	24.1	22	100.0	34	-	-	-	-	97.9	14	99.9	18	99.2	22	-	-	Autotransporter of unknc	IPR005546:	Yes	ACSGSSTV	1DAB;Cla	n.d.	428-554	L3 (165-2
YP_4571Serine proteaErythrobacMKTRKYLVCVATNAFFI	-	98.4	88	98.4	16	95.7	14	-	-	87.2	19	95.2	22	96.1	21	95.9	21	97.6	20	-	-	-	-	34.9	21	93.6	18	15.8	17	-	-	Protease AutoTransporte	IPR005546:	Yes	NDDNDASI	4O03;Cla	n.d.	180-297	n.d.
YP_0014Outer membrSerratia prMKRLAIAIIAALSDAGLI	-	96.4	95	77.1	15	-	-	-	-	52.8	13	-	-	-	-	-	-	96.4	14	-	-	-	-	5.7	8	36.3	11	2.8	15	-	-	Autotransporter of unknc	IPR005546:	Yes	VESYEVRN	1LML;Cla	IPR011050	428-554	n.d.
YP_2614Outer membrPseudomoMQRLVRSTIWNIRGDS	-	100.0	46	98.6	30	96.7	24	-	-	96.3	22	98.3	29	97.9	31	98.0	28	100.0	33	-	-	97.6	32	99.5	28	100.0	35	99.9	32	-	-	Autotransporter of unknc	IPR005546:	Yes	AEYGQIDL	1DAB;Cla	n.d.	180-297	L1 (88-15
YP_0041Outer membrMesorhizoMRVPREAYLFINGNVT	-	99.9	68	91.9	14	-	-	-	-	-	-	66.8	15	-	-	-	-	99.9	17	-	-	47.5	21	98.6	18	99.9	18	99.2	11	-	-	Protease / Peptidase M1	IPR005546:	Yes	CSAIASGA	n.d.	n.d.	91-203	L1 (415-5
YP_2607Outer membrPseudomoMRVSGAQNRGDSQLT	-	100.0	165	92.6	22	84.1	24	-	-	32.3	20	96.9	41	84.6	30	86.0	28	100.0	32	-	-	95.9	19	98.8	25	100.0	45	99.8	35	-	-	Autotransporter of unknc	IPR005546:	Yes	ADGAAIEL	1DAB;Cla	n.d.	180-297	L1 (139-1
YP_0027Adhesin aidaLaribacterMKPVAVRPIAGSEVAG	-	100.0	67	99.5	24	98.6	15	-	-	98.7	18	99.6	28	99.5	22	99.4	12	99.9	25	-	-	74.6	20	98.8	18	100.0	29	99.7	25	-	-	Self-Associating AutoTra	IPR030930:	Yes	DFTTPVTS	1DAB;Cla	n.d.	473-603	L1 (418-5
YP_5766Outer membrNitrobacteMYRVRTSPIGTITNAGL	-	100.0	152	99.6	37	99.0	19	-	-	98.8	16	99.4	28	99.3	25	99.1	19	100.0	38	-	-	21.7	29	98.5	16	99.9	26	99.3	26	-	-	Autotransporter of unknc	IPR005546:	Yes	AEVSGLGH	1DAB;Cla	n.d.	180-297	L1 (117-1
YP_0014Outer membrSerratia prMASFAPSFFWTVTHDS	-	100.0	51	96.5	21	34.7	17	-	-	76.6	24	96.6	37	95.4	31	96.6	31	99.9	24	-	-	97.8	26	99.4	27	100.0	38	99.8	38	-	-	Autotransporter of unknc	IPR005546:	Yes	ASPWRVVI	1DAB;Cla	n.d.	428-554	L1 (109-2
YP_0029AutotransporEdwardsieMHKFSFPALLNHAGNL	-	100.0	48	100.0	30	99.9	17	93.3	20	99.8	14	99.9	29	99.9	24	99.9	20	100.0	30	-	-	88.0	14	98.8	14	100.0	28	99.7	25	-	-	Phosphatase/Hydrolase	IPR005546:	Yes	QQDNALRI	1DAB;Cla	n.d.	473-603	n.d.
YP_0033Outer membrCitrobacteMYINMKFISILVVDTSLS	-	100.0	45	99.3	44	96.9	18	-	-	95.9	15	98.2	23	97.8	29	98.1	20	100.0	40	-	-	66.9	17	98.7	16	100.0	23	99.7	25	-	-	Autotransporter of unknc	IPR005546:	Yes	STLYFNSSI	1DAB;Cla	n.d.	265-383	L3 (148-2
YP_0039Outer membrAchromobMKCAEHEDSNLTNLVI	-	100.0	56	98.4	31	97.2	24	-	-	96.2	17	98.4	29	97.9	29	97.9	25	100.0	33	-	-	97.0	26	99.0	22	100.0	36	99.8	30	-	-	Autotransporter of unknc	IPR005546:	Yes	EIIPQGTYII	1DAB;Cla	n.d.	91-203	L1 (393-4
YP_0036Outer membrBurkholderMRPCPIRQ/ASVVNAG	-	100.0	105	96.5	23	31.5	16	-	-	-	-	87.0	25	-	-	33.3	21	100.0	25	-	-	80.3	18	98.7	15	99.9	21	99.5	26	-	-	Autotransporter of unknc	IPR005546:	Yes	QQNLHIDG	1DAB;Cla	n.d.	200-322	L1 (76-14
YP_0041Outer membrRhodopserMRVYRDTILGSLDLRG	-	97.6	99	97.6	17.0	88.4	15	-	-	88.3	11	93.5	13	94.8	17	89.8	15	96.8	18	-	-	-	-	75.4	12	95.0	15	31.0	17	-	-	Phosphata Phosphatase	IPR005546:	Yes	QQQPSCTC	2AKC;Cla	n.d.	91-203	n.d.
NP_882CBrka autotranBordetellaMYLDRFRQVVNADSR	-	100.0	60	99.3	21	98.9	21	-	-	98.1	22	99.6	44	99.3	28	99.3	30	100.0	19	-	-	97.7	25	99.3	25	100.0	42	99.9	27	-	-	Autotransporter of unknc	IPR005546:	Yes	AAQAPQPF	1DAB;Cla	n.d.	n.d.	L1 (117-1
YP_0030Outer membrPectobactMTAAQLFSLGNASVSG	-	96.2	130	95.5	16.0	39.2	14	-	-	27.9	14	84.9	17	89.6	22	-	-	96.2	17	-	-	8.8	12	35.0	15	94.2	20	54.0	19	-	-	Protease / Peptidase S8	IPR005546:	Yes	GGGETASV	1R6V;Cla	n.d.	322-432	n.d.
YP_0033Outer membrCitrobacteMDHYHFSQGQLTNNG	-	100.0	77	98.8	39	96.2	17	-	-	97.2	10	97.4	27	97.1	33	97.5	26	100.0	41	-	-	81.9	21	99.0	17	100.0	25	99.7	31	-	-	Autotransporter of unknc	IPR005546:	Yes	ANSNVIVVI	1DAB;Cla	n.d.	265-383	L1 (184-2
YP_0039Outer membrEnterobactMQKKYLVMFTVNGAL	-	100.0	65	97.3	35	76.6	14	-	-	74.0	13	90.5	21	82.5	19	90.1	16	100.0	39	-	-	91.1	15	98.8	12	99.9	24	99.7	27	-	-	Autotransporter of unknc	IPR005546:	Yes	AQCTTASS	1DAB;Cla	n.d.	91-203	n.d.
YP_0042Outer membrRahnella sMGNNPEKWQWLMTDS	-	100.0	80	98.6	20	97.0	21	-	-	96.4	21	98.4	31	98.0	25	98.2	25	100.0	25	-	-	97.8	24	99.6	22	100.0	31	99.9	27	-	-	Autotransporter of unknc	IPR005546:	Yes	TDISDKDEI	1DAB;Cla	n.d.	91-203	L1 (165-2
YP_3741Outer membrChlorobiumMKRRILRLILGRDRIAG	-	99.4	40	99.4	20	98.7	14	30.6	16	98.7	17	98.8	29	99.0	22	98.6	21	97.6	24	-	-	-	-	-	-	71.2	36	-	-	-	-	Autotransporter of unknc	IPR005546:	Yes	KEFVVSES	n.d.	IPR011050	180-297	L3 (88-18
YP_0040Hypothetical Neisseria lMIKITYSKNCDSMLKEL	-	99.8	61	99.7	14	99.5	27	98.0	22	99.3	21	99.8	22	99.8	26	99.7	21	97.8	17	-	-	-	-	39.4	25	98.3	24	93.9	26	-	-	Autotransporter of unknc	IPR005546:	Yes	ADVNKGTI	1DAB;Cla	n.d.	91-203	n.d.
YP_0013Outer membrOchrobactMKKLWLASFVLQGNV	-	100.0	79	99.5	38	97.7	14	-	-	97.4	12	98.7	23	98.6	29	98.5	19	100.0	39	-	-	91.1	18	99.0	16	100.0	25	99.7	30	-	-	Autotransporter of unknc	IPR005546:	Yes	APCASNSN	1DAB;Cla	n.d.	428-554	L1 (174-2
YP_0033Outer membrSebaldellaMKIKNTLKNDLNNSGK	-	100.0	81	99.6	32	98.6	21	-	-	98.2	18	99.2	25	99.0	28	99.1	19	100.0	33	-	-	94.7	14	99.0	19	100.0	26	99.7	27	-	-	Autotransporter of unknc	IPR005546:	Yes	KNIYEVSG	1DAB;Cla	n.d.	265-383	L1 (305-3
YP_0011Outer membrEnterobactMSCMSNKALTIEGDLT	-	100.0	102	96.6	18	-	-	-	-	31.6	14	38.3	29	-	-	50.7	22	100.0	39	-	-	72.5	23	98.8	20	100.0	28	99.7	29	-	-	Autotransporter of unknc	IPR005546:	Yes	TTVDTNTT	1DAB;Cla	n.d.	428-554	L1 (158-2
YP_0033Adhesin autoCitrobacteMSVTVTDRWNITGND	-	100.0	147	99.3	25	98.7	22	84.9	24	98.6	21	99.4	40	99.5	35	99.4	32	99.9	22	-	-	98.2	30	99.6	23	100.0	41	99.9	35	-	-	Autotransporter of unknc	IPR005546:	Yes	VSVSGTSV	1DAB;Cla	n.d.	265-383	L1 (253-3
YP_0029Pertactin CandidatusMLIRVQKKRVSRFGKV	-	100.0	187	99.0	20	98.1	19	-	-	96.4	14	99.5	32	99.3	25	99.5	25	99.9	23	-	-	96.0	21	98.9	14	100.0	34	99.8	22	-	-	Autotransporter of unknc	IPR005546:	Yes	SFTETVAE	1DAB;Cla	n.d.	473-603	T4 (95-15
YP_2078Hypothetical Neisseria gMNQEGITAHWTMTGQ	-	98.8	81	98.8	21	97.9	21	-	-	97.9	17	98.6	26	98.7	19	98.3	17	94.2	21	-	-	-	-	86.3	30	97.7	32	79.8	25	-	-	Autotransporter of unknc	IPR005546:	Yes	NNKITVEN	1DAB;Cla	n.d.	428-554	V1 (265-3
YP_0018Outer membrBeijerinckiMKPNVPSAINGNVTDV	-	98.8	80	98.8	24.0	91.8	12	-	-	91.6	8	96.3	19	96.0	17	95.8	13	97.5	28	-	-	-	-	-	-	94.0	19	-	-	-	-	Autotransporter of unknc	IPR005546:	Yes	QTHDIDRA	1DAB;Cla	IPR011050	428-554	TS (443-4
YP_0042Outer membrRahnella sMINIDLPTSFSATVTGD	-	100.0	109	99.6	39	98.1	14	-	-	98.0	19	98.7	24	98.7	33	98.5	19	100.0	39	-	-	66.9	26	98.4	19	99.9	29	99.6	31	-	-	Autotransporter of unknc	IPR005546:	Yes	AGLNVTIGI	1DAB;Cla	n.d.	91-203	n.d.
YP_0039Outer membrEnterobactMSNKKDVSIMTNGGVI	-	100.0	86	46.5	32	-	-	-	-	-	-	-	-	-	-	-	-	100.0	36	-	-	-	-	98.8	28	100.0	24	99.6	31	-	-	Autotransporter of unknc	IPR005546:	Yes	VINNSETIID	1DAB;Cla	n.d.	91-203	L1 (72-164
YP_0039Outer membrAchromobMTTPKSRQIGASVNNS	-	96.4	114	40.5	18	-	-	-	-	-	-	-	-	-	-	-	-	96.4	18	-	-	19.6	16	-	-	69.8	18	23.1	11	-	-	Protease / Peptidase S8	IPR005546:	Yes	GGGGGGG	2OXA;Cla	n.d.	91-203	n.d.
YP_0042Outer membrRahnella sMPAFNTCKISVLGDLSI	-	100.0	46	99.3	30	98.6	30	88.3	19	98.4	22	99.0	35	98.7	33	99.0	28	100.0	32	-	-	96.4	22	99.1	22	100.0	35	99.7	34	-	-	Autotransporter of unknc	IPR005546:	Yes	ANYTLPLLL	1DAB;Cla	n.d.	91-203	L1 (268-3
YP_0013AutotransporJanthinobaMKQKLTYTLSTINNLSL	-	100.0	50	99.4	34	98.5	19	-	-	98.1	19	99.0	24	98.8	29	98.8	22	100.0	33	-	-	95.9	19	99.2	21	100.0	24	99.8	28	-	-	Autotransporter of unknc	IPR005546:	Yes	SDGCTVNC	1DAB;Cla	n.d.	428-554	n.d.
NP_9015Abc transporChromobaMRTQAVSRISDVGQLA	-	100.0	47	96.8	25	89.0	14	-	-	92.9	19	97.5	36	97.4	33	97.6	30	100.0	24	-	-	96.1	20	98.9	19	100.0	39	99.9	32	-	-	Autotransporter of unknc	IPR005546:	Yes	GAVGAPLR	1DAB;Cla	n.d.	428-554	L1 (314-3
YP_0023Iga1 proteaseNeisseria rMKTKRFKINWHLTGDS	-	100.0	541	98.8	20	98.8	24	95.9	22	98.3	19	99.4	26	99.4	39	99.6	47	99.9	21	-	-	97.6	22	99.1	22	100.0	26	99.8	35	-	-	Protease / Peptidase S6	IPR005546:	Yes	ALVRDDVD	1DAB;Cla	n.d.	473-603	T6 (777-8
YP_0030Iga-specific sNeisseria rMKTKRFKINWHLTGDS	-	100.0	620	98.3	20	98.1	25	87.1	18	97.7	21	99.1	26	99.1	39	99.3	47	99.9	20	-	-	98.2	24	99.3	21	100.0	27	99.9	37	-	-	Protease / Peptidase S6	IPR005546:	Yes	ALVRDDVD	1DAB;Cla	n.d.	265-383	T6 (789-8
YP_2074Iga-specific sNeisseria gMKAKRFKINWHLTGDS	-	100.0	371	94.6	20	89.1	23	-	-	77.5	19	98.0	27	98.1	35	98.6	47	99.9	20	-	-	97.4	25	98.9	20	100.0	27	99.9	37	-	-	Protease / Peptidase S6	IPR005546:	Yes	ALVRDDVD	1DAB;Cla	n.d.	91-203	T6 (788-8
YP_0020Iga-specific nNeisseria gMKAKRFKINWHLTGDS	-	100.0	371	96.0	21	94.4	23	-	-	91.2	17	98.3	27	98.2	38	98.7	46	99.9	19	-	-	97.5	23	99.3	20	100.0	27	99.8	34	-	-	Protease / Peptidase S6	IPR005546:	Yes	ALVRDDVD	1DAB;Cla	n.d.	132-263	T6 (788-8
YP_0030Serine-type pNeisseria rMRFTHTTPFSSHTGAL	-	100.0	208	99.3	23	98.4	19	68.4	24	97.9	22	99.5	31	99.6	45	99.4	31	100.0	24	-	-	97.0	20	98.8	19	100.0	33	99.9	46	-	-	Protease / Peptidase S6	IPR005546:	Yes	SIVRNDVD	1DAB;Cla	n.d.	265-383	n.d.
NP_2749Adhesion andNeisseria rMKTTDKRTTGNLNLDN	-	100.0	152	98.9	23	97.0	17	59.2	19	95.3	18	99.1	37	99.4	75	99.0	31	100.0	20	-	-	93.2	19	98.3	23	100.0	29	100.0	51	-	-	Protease / Peptidase S6	IPR005546:	Yes	GHTYFGIN	1DAB;Cla	n.d.	n.d.	T3 (798-8
NP_8384Serine proteaShigella flеMNKIYSLKYQSALQSL	-	100.0	75	99.3	13	99.4	46	98.9	33	99.3	43	99.0	22	98.9	21	99.0	19	99.9	10	-	-	99.7	36	100.0	44	99.9	26	99.4	22	-	-	Protease / Peptidase S6	IPR005546:	Yes	AMLDINNIV	1DAB;Cla	n.d.	n.d.	T3 (627-6
NP_7087Serine proteaShigella flеMNKIYSLKYQSALQSL	-	99.9	75	99.3	13	99.4	46	98.9	33	99.3	43	99.0	22	98.9	21	99.0	19	99.9	11	-	-	99.7	36	99.9	43	99.9	24	99.3	18	-	-	Protease / Peptidase S6	IPR005546:	Yes	AMLDINNIV	1DAB;Cla	n.d.	n.d.	T3 (627-6
YP_0024Serine proteaEscherichiMNKIYSLKYQSALQSL	-	99.9	75	99.3	13	99.4	46	98.9	33	99.3	43	99.0	22	98.9	21	99.0	19	99.9	11	-	-	99.7	36	99.9	43	99.9	24	99.3	18	-	-	Protease / Peptidase S6	IPR005546:	Yes	AMLDINNIV	1DAB;Cla	n.d.	473-603	n.d.
YP_3120Serine proteaShigella scMNKIYSLKYQSALQSL	-	100.0	72	99.3	13	99.5	46	99.0	33	99.4	43	99.1	22	99.0	21	99.1	19	99.9	10	-	-	99.7	36	100.0	44	99.9	26	99.4	22	-	-	Protease / Peptidase S6	IPR005546:	Yes	AMLDINNIV	1DAB;Cla	n.d.	180-297	T3 (627-6
YP_0018Serine proteaShigella bcMNKIYSLKYRVNGQSA	-	100.0	74	99.2	14	99.4	46	98.9	36	99.3	43	98.8	22	98.8	21	99.0	19	99.9	13	-	-	99.7	35	100.0	44	99.9	25	99.2	21	-	-	Protease / Peptidase S6	IPR005546:	Yes	AMLDINNIV	1DAB;Cla	n.d.	132-263	T3 (627-6
YP_0032Predicted serEscherichiMLILCCLGILDSGLNTLI	-	100.0	37	98.5	11	99.3	50	98.9	36	99.4	50	98.5	23	98.5	22	98.6	27	97.5	14	-	-	99.9	36	100.0	50	97.6	28	97.1	28	-	-	Protease / Peptidase S6	IPR005546:	Yes	SRMDASNF	1DAB;Cla	n.d.	265-383	T3 (607-6
YP_0030Putative exopEscherichiMNKIYSLKYDSALKTLK	-	99.9	77	98.3	11	99.3	100	98.8	46	99.2	59	98.4	25	98.0	21	98.2	25	99.9	14	-	-	99.7	44	99.9	60	99.9	28	99.4	22	-	-	Protease / Peptidase S6	IPR005546:	Yes	SQMDISNF	1DAB;Cla	n.d.	265-383	T3 (620-6
NP_0526Espp (plasmiEscherichiMNKIYSLKYDSALKTLK	-	100.0	77	98.3	11	99.3	100	98.8	46	99.2	59	98.4	25	98.0	21	98.2	25	99.9	14	-	-	99.7	44	99.9	60	100.0	28	99.4	22	-	-	Protease / Peptidase S6	IPR005546:	Yes	SQMDISNF	1DAB;Cla	n.d.	n.d.	n.d.
YP_3255Putative exopEscherichiMNKIYSLKYDSALKTLK	-	99.9	77	98.3	11	99.3	100	98.8	46	99.2	59	98.4	25	98.0	21	98.2	25	99.9	14	-	-	99.7	44	99.9	60	99.9	28	99.4	22	-	-	Protease / Peptidase S6	IPR005546:	Yes	SQMDISNF	1DAB;Cla	n.d.	180-297	n.d.
YP_0022ImmunoglobuEscherichiMNKIYSLKYDSALKTLK	-	99.9	77	98.3	11	99.3	100	98.8	46	99.2	59	98.4	25	98.0	21	98.2	25	99.9	14	-	-	99.7	44	99.9	60	99.9	28	99.4	22	-	-	Protease / Peptidase S6	IPR005546:	Yes	SQMDISNF	1DAB;Cla	n.d.	132-263	n.d.
YP_0024Serine proteaEscherichiMREYMNKIYWHLNSQS	-	100.0	75	98.4	13	99.3	54	98.9	48	99.5	59	98.4	27	98.6	16	98.5	25	99.9	14	-	-	99.8	47	100.0	57	99.9	30	99.5	20	-	-	Protease / Peptidase S6	IPR005546:	Yes	ANIDISNVW	1DAB;Cla	n.d.	473-603	T3 (645-6
WP_014Plasmid encoEscherichiMNKIYSIKYSNSELKKLI	-	100.0	374	98.3	11	97.9	14	98.4	95	99.3	100.0	98.4	25	98.3	33	99.1	32	97.9	23		-	98.1	31	100.0	43	98.0	27	98.1	24		-	Protease / Peptidase S6	IPR005546:	Yes	ANMDISKA	1DAB;Cla	n.d.	428-554	n.d.
YP_0023 serine proteaEscherichiMNKIYALKYNSTIGTLK	-	100.0	84	98.4	13	99.1	49	98.5	38	99.1	47	98.5	26	98.5	18	98.5	23	99.9	14	-	-	99.7	36	100.0	49	99.9	25	99.5	22	-	-	Protease / Peptidase S6	IPR005546:	Yes	DYLDLAQN	1DAB;Cla	n.d.	473-603	L1 (691-7
YP_0023Serine proteaEscherichiMNRIYSLRNNSSVKHL	-	100.0	154	97.5	14	98.9	45	98.5	36	99.0	47	97.3	26	97.2	14	97.4	21	99.8	13	-	-	99.8	36	100.0	47	99.8	28	99.3	16	-	-	Protease / Peptidase S6	IPR005546:	Yes	VVGNTLPW	1DAB;Cla	n.d.	473-603	T3 (636-6
YP_0017Outer membrEscherichiMNFVYTLKLLGNLTLKN	-	99.9	73	43.7	-	92.5	36	95.8	26	95.0	34	-	-	-	-	-	-	93.2	14	-	-	99.8	27	99.9	34	72.5	19	59.9	7	-	-	Protease / Peptidase S6	IPR005546:	Yes	IVGMDIPYC	1DAB;Cla	n.d.	428-554	T3 (690-7
YP_0033Adhesin autoCitrobacteMNKIYAIKKNWQADGG	-	100.0	74	98.0	9	99.0	36	98.6	33	99.0	41	98.3	24	98.5	18	98.1	23	99.8	10	-	-	99.8	34	100.0	41	99.9	25	99.5	18	-	-	Protease / Peptidase S6	IPR005546:	Yes	DVGNILPYC	1DAB;Cla	n.d.	265-383	T3 (636-7
NP_7523Hemoglobin pEscherichiMNKIYALKYNSTAGNM	-	100.0	73	97.7	20	98.8	42	98.4	95	99.0	41	97.1	26	97.7	24	97.3	26	97.9	23	-	-	99.9	97	100.0	41	98.0	27	97.9	24	-	-	Protease / Peptidase S6	IPR005546:	Yes	TVGGNNPY	1DAB;Cla	n.d.	n.d.	n.d.
YP_0023Vacuolating aEscherichiMNKIYALKYNSTAGNM	-	100.0	73	97.8	20	98.9	42	98.4	95	99.1	41	96.8	25	97.6	24	97.1	27	97.9	23	-	-	99.9	97	100.0	41	98.0	27	97.9	24	-	-	Protease / Peptidase S6	IPR005546:	Yes	TVGGNNPY	1DAB;Cla	n.d.	473-603	n.d.
YP_8514Vacuolating aEscherichiMNKIYALKYNSTAGNM	-	100.0	73	97.8	20	98.9	42	98.4	95	99.1	41	96.8	25	97.6	24	97.1	27	97.9	23	-	-	99.9	97	100.0	41	98.0	27	97.9	24	-	-	Protease / Peptidase S6	IPR005546:	Yes	TVGGNNPY	1DAB;Cla	n.d.	345-479	n.d.
YP_0029Secreted proEdwardsieMNKVYSLRYDSTLGSLI	-	100.0	72	96.2	21	98.5	39	98.1	34	98.5	32	95.5	20	95.6	16	96.1	19	97.7	22	-	-	99.9	34	100.0	31	97.8	30	97.6	21	-	-	Protease / Peptidase S6	IPR005546:	Yes	IVSDEIPYS	1DAB;Cla	n.d.	473-603	T3 (704-7
NP_8582Extracellular Shigella flеMNKIYYLKYSEAGALK	-	99.9	77	97.9	14	98.9	46	98.2	36	98.9	42	98.1	26	98.4	18	98.0	24	99.9	15	-	-	99.6	35	99.9	44	99.9	24	99.5	18	-	-	Protease / Peptidase S6	IPR005546:	Yes	TVSAEIPYC	1DAB;Cla	n.d.	n.d.	T3 (698-7
YP_0014Secreted serEscherichiMNKVFSLKYRNSEIKEL	-	99.9	77	98.3	12	99.1	41	98.2	35	99.1	44	98.5	26	98.6	17	98.4	20	99.8	13	-	-	99.5	32	99.9	46	99.8	25	99.4	17	-	-	Protease / Peptidase S6	IPR005546:	Yes	TVNADISYC	1DAB;Cla	n.d.	428-554	T3 (700-7
YP_0032Secreted autEscherichiMNKIYALKYDSSLRSLI	-	100.0	74	98.5	9	99.3	65	98.1	45	99.1	55	98.6	24	98.2	20	98.5	23	99.9	9	-	-	99.7	44	100.0	55	99.9	27	99.4	19	-	-	Protease / Peptidase S6	IPR005546:	Yes	TVSAEIPYC	1DAB;Cla	n.d.	265-383	T3 (692-7
NP_8384Serine proteaShigella flеMNKVYSLKYDSAIHSLT	-	100.0	79	97.9	13	99.0	40	98.7	46	99.2	39	97.5	25	97.7	24	97.6	27	97.7	14	-	-	99.9	45	100.0	39	97.8	27	97.7	25	-	-	Protease / Peptidase S6	IPR005546:	Yes	IVRSDIAYQ	1DAB;Cla	n.d.	n.d.	n.d.
YP_0013Outer membrSinorhizobMPDISPGRTTVYNAGII	-	100.0	141	98.8	33	95.1	21	-	-	91.4	13	97.5	26	97.1	28	96.5	19	100.0	34	-	-	82.0	24	98.9	19	100.0	19	99.6	30	-	-	Autotransporter of unknc	IPR005546:	Yes	SEARAACV	1DAB;Cla	IPR011050	428-554	L3 (277-3
NP_2466Hypothetical PasteurellaMTHLKKSSLYELNGKF	-	100.0	47	99.7	28	98.8	11	-	-	98.3	12	99.3	19	99.0	21	99.1	18	100.0	28	-	-	91.8	13	98.9	12	100.0	17	99.7	21	-	-	Autotransporter of unknc	IPR005546:	Yes	VNTTSCGE	1DAB;Cla	IPR011050;IPF	n.d.	n.d.
YP_0010AutotransporYersinia eMHKIYRFIKTTGKVNNA	-	100.0	63	99.3	29	98.3	11	-	-	97.7	11	98.5	19	98.3	21	98.4	17	100.0	30	-	-	95.2	21	99.1	13	100.0	19	99.8	21	-	-	Autotransporter of unknc	IPR005546:	Yes	AETALNTQ	1DAB;Cla	n.d.	428-554	n.d.
YP_0015Hypothetical SalmonellaMKKLLALLVFKFNNDL	-	100.0	52	98.1	40	79.2	20	-	-	91.5	18	96.1	22	93.1	28	94.7	15	100.0	40	-	-	86.6	18	99.1	16	100.0	22	99.8	30	-	-	Autotransporter of unknc	IPR005546:	Yes	TYIDTLGNE	1DAB;Cla	n.d.	428-554	L1 (154-2