Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic *Escherichia coli*

Danilo Gomes Moriel[a], Isabella Bertoldi[a], Angela Spagnuolo[a], Sara Marchi[a], Roberto Rosini[a], Barbara Nesta[a], Ilaria Pastorello[a], Vanja A. Mariani Corea[a], Giulia Torricelli[a], Elena Cartocci[a], Silvana Savino[a], Maria Scarselli[a], Ulrich Dobrindt[b], Jörg Hacker[b,c], Hervé Tettelin[d,2], Luke J. Tallon[d,2], Steven Sullivan[d,3], Lothar H. Wieler[e], Christa Ewers[e], Derek Pickard[f], Gordon Dougan[f], Maria Rita Fontana[a], Rino Rappuoli[a,1], Mariagrazia Pizza[a], and Laura Serino[a,1]

[a]Novartis Vaccines and Diagnostics, 53100 Siena, Italy; [b]Institut für Molekulare Infektionsbiologie, Universität Würzburg, D-97080 Würzburg, Germany; [c]German Academy of Sciences Leopoldina, D-06108 Halle, Germany; [d]J. Craig Venter Institute, Rockville, MD 20850; [e]Institut für Mikrobiologie und Tierseuchen, Freie Universität Berlin, D-10061 Berlin, Germany; and [f]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK

Contributed by Rino Rappuoli, December 29, 2009 (sent for review December 21, 2009)

Extraintestinal pathogenic *Escherichia coli* (ExPEC) are a common cause of disease in both mammals and birds. A vaccine to prevent such infections would be desirable given the increasing antibiotic resistance of these bacteria. We have determined the genome sequence of ExPEC IHE3034 (ST95) isolated from a case of neonatal meningitis and compared this to available genome sequences of other ExPEC strains and a few nonpathogenic *E. coli*. We found 19 genomic islands present in the genome of IHE3034, which are absent in the nonpathogenic *E. coli* isolates. By using subtractive reverse vaccinology we identified 230 antigens present in ExPEC but absent (or present with low similarity) in nonpathogenic strains. Nine antigens were protective in a mouse challenge model. Some of them were also present in other pathogenic non-ExPEC strains, suggesting that a broadly protective *E. coli* vaccine may be possible. The gene encoding the most protective antigen was detected in most of the *E. coli* isolates, highly conserved in sequence and found to be exported by a type II secretion system which seems to be nonfunctional in nonpathogenic strains.

urinary tract infections | sepsis

Escherichia coli is a common colonizer of the human gastrointestinal tract, often appearing soon after birth and persisting for decades. Although *E. coli* are largely regarded as commensals, some isolates have the potential to also cause disease. Pathogenic *E. coli* can be divided into three major subgroups depending on their pathogenic traits: commensals or nonpathogenic, pathogenic causing intestinal infections, and extraintestinal pathogenic *E. coli* (ExPEC). The ExPEC group includes human and animals pathogens causing urinary tract infections (UPEC), and others causing neonatal meningitis (NMEC) and septicemia. Other ExPEC are pathogenic to avian species (APEC). ExPEC strains account for over 80% of urinary tract infection cases and are the second major cause of neonatal meningitis and sepsis. The intestinal pathogenic *E. coli* group includes many pathotypes such as enterotoxigenic (ETEC), enteropathogenic, enterohemorrhagic (EHEC), enteroinvasive, adherent invasive, and diffusely adherent *E. coli*, all causing infections to the human intestinal tract (1).

The unique ability to colonize different sites and to cause a heterogeneous group of disorders is potentially due, at least in part, to the versatile genome of *E. coli*, which are able to remodel their genetic repertoire by acquiring and losing virulence attributes. This plasticity is reflected in the size of *E. coli* genomes that vary from 4.64 Mb for the nonpathogenic *E. coli* K12 isolate MG1655 (2) to 5.70 Mb for the enterohemorrhagic strain 11368 (3). Differences in genome size are primarily due to the insertion or deletion of a few large chromosomal regions. Horizontal gene transfer plays a major role in the acquisition or loss of DNA and usually contributes to the evolution and adaptation of *E. coli* to different niches (4).

The development of an efficacious ExPEC vaccine would have significant public health and economic impacts, considering the increasing antibiotic resistance among ExPEC strains and the associated mortality, morbidity, and lost productivity (5). Conventional vaccinology approaches have mainly relied on whole-cell, single antigens or polysaccharide-based approaches, but these have been largely unsuccessful in providing a highly immunogenic, safe, and cross-protective vaccine against ExPEC strains (6).

Today, multiple genomic sequences of *E. coli* have been completely or partially determined providing an opportunity to use a "reverse vaccinology" approach to identify vaccine candidates (7).

Because most of the ExPEC genomes available up to now are from UPEC strains, we decided to sequence a neonatal meningitis-associated K1 *E. coli*, IHE3034 (8). The comparison of this genome with those of pathogenic and nonpathogenic *E. coli* strains facilitated the identification of unique genomic features and of broadly protective vaccine candidates.

Results

IHE3034 Genome Features and Genome Comparison. We determined the complete chromosome sequence of the *E. coli* K1 strain IHE3034 (O18:K1:H7), belonging to the EcoR B2 group and to multilocus sequence type 95 (ST95) and isolated in Finland in 1976 from a case of neonatal meningitis (8). The assembly of IHE3034 genome sequence (GenBank accession no. CP001969) revealed a 5,108,383-bp circular chromosomal sequence consisting of 4,805 annotated protein-coding regions, 21 rRNAs, 96 tRNAs and a G + C% content of 50.7%.

To identify genomic features restricted to IHE3034, we aligned this genome with those of other sequenced strains. The IHE3034

Author contributions: D.G.M., R. Rappuoli, M.P., and L.S. designed research; D.G.M., I.B., A.S., S.M., R. Rosini, B.N., I.P., V.A.M.C., G.T., E.C., H.T., C.E., and M.R.F. performed research; U.D. and C.E. contributed new reagents/analytic tools; S. Savino, M.S., U.D., J.H., L.H.W., D. P., G.D., L.J.T., S. Sullivan, M.R.F., and L.S. analyzed data; and D.G.M., R. Rappuoli, M.P., and L.S. wrote the paper.

Conflict of interest statement: R. Rappuoli is a full-time employee of Novartis Vaccines.

Freely available online through the PNAS open access option.

Data deposition: The complete sequence and annotation of the IHE3034 strain has been deposited in the GenBank database (accession no. CP001969).

[1]To whom correspondence may be addressed: E-mail: rino.rappuoli@novartis.com or laura.serino@novartis.com.

[2]Current address: Institute for Genome Sciences, School of Medicine, University of Maryland, Baltimore, MD 21201.

[3]School of Medicine, New York University, New York, NY 10010.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.0915077107/-/DCSupplemental.

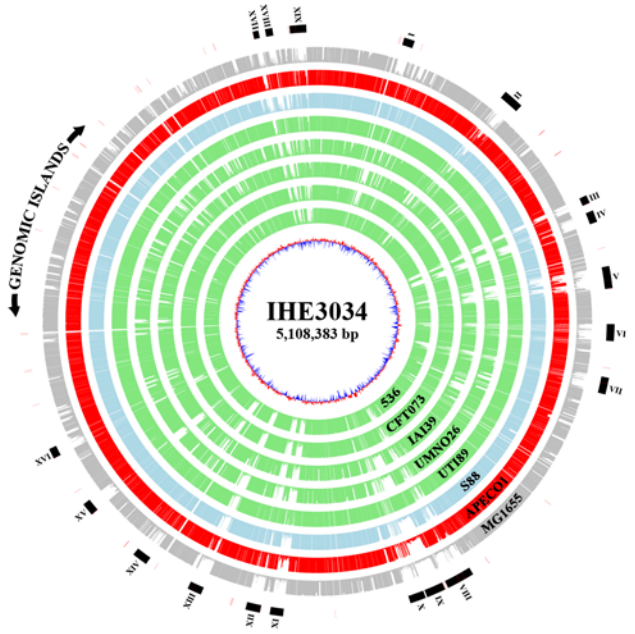

Fig. 1. Comparative genome analysis between IHE3034 and other ExPEC strains. Comparison between IHE3034 ORFs with the nonpathogenic *E. coli* strain MG1655 (external gray) and other seven ExPEC strains. Nineteen genomic islands (black boxes in the outside, identified by roman numerals), specific to ExPEC strains, and not present in the nonpathogenic strain MG1655 are shown. The green circles represent UPEC strains, the light blue is a human K1 strain, and the red is an avian K1 strain.

genome sequence was initially compared to that of nonpathogenic *E. coli* K12 strain MG1655 (2), shown in gray in the outer circle of Fig. 1. This comparison identified 19 genomic islands (>10 Kb) absent in the MG1655 genome (Fig. 1 and Table S1). As shown in Table S1, the size of the IHE3034-specific islands varies from 16 to 78 Kb and altogether account for 777 Kb. Many of the islands have the typical features of pathogenicity islands (PAIs), i.e., they are associated with tRNAs genes and have a G + C% content different from the core genome. The genomic islands that we consider meet the definition of PAIs are I, V, IX, XV, and XVI. The 19 genomic islands harbor genes coding for many candidate virulence factors, such as type VI secretion systems (T6SS, islands I and XV) (9); S-fimbrial determinants (island V) (10); autotransporter proteins (island V) (11); the siderophore receptor IroN (island V) (12); iron transporters, SitABCD (island VI) (13); the yersiniabactin gene cluster (island VIII) (14); cytolethal distending toxin (island VIII) (15); a colibactin gene cluster (island IX) (16); IbrA and IbrB proteins (island X) (17); enterohemolysin 1 (island XIII) (18); a type II secretion system (T2SS, island XVI) (19); the K1 capsule locus (island XVI) (20); and IbeA and IbeT (island XVIII) (21, 22). We then compared the IHE3034 sequence to the genomes of other ExPEC strains such as APECO1 (23), S88 (24), UTI89 (25), UMNO26 (24), IAI39 (24), CFT073 (26), and 536 (27) (Fig. 1: red, APEC; light blue, NMEC; green, UPEC). We found that only part of island VI and VIII, encoding a *Salmonella* iron transporter and the yersiniabactin gene cluster respectively, are present and highly conserved in all strains, confirming the high plasticity of the genomes and the high variability of the *E. coli* causing extraintestinal infections (27–29).

Selection of Potential Vaccine Candidates by Subtractive Reverse Vaccinology. To identify potential vaccine candidates against ExPEC, we applied a "subtractive reverse vaccinology" approach. Briefly, antigens predicted to be surface associated or secreted and with no more than three transmembrane domains (TMDs) were selected by bioinformatic analysis of the IHE3034, 536, and CFT073 genomes. The presence (and the level of similarity) of these antigens in the nonpathogenic strains MG1655, DH10B, and W3110 were used as exclusion criteria. By this approach, we were able to identify 230 potential antigens (black bars in Fig. S1), which were then expressed as His-tagged proteins, purified and tested for protection in a sepsis mouse model. In order to improve the solubility or expression levels of some of the antigens, the genes were expressed as smaller peptides, yielding a total of 270 candidates. Of these, 220 were successfully purified, 69 as soluble and 151 as insoluble proteins. In vivo protection was evaluated in CD1 mice by subcutaneous injection with three doses of 20 μg for each purified recombinant antigen, in Freund's adjuvant at days 1, 21, and 35. Two weeks after the last dose, mice were challenged by intraperitoneal injection of the *E. coli* pathogenic strains (1×10^7 CFU/mL for strains IHE3034, 3×10^7 CFU/mL CFT073) or intravenous (3×10^7 CFU/mL for strain 536). After challenge, animal survival was followed for 4 d and bacteremia was measured at 20 h. We found a good correlation between mortality and the number of bacteria in the blood. By using this approach, we identified nine potential vaccine candidates, able to induce significant level of protection compared to the control group. The vaccine candidate protection level following an homologous lethal challenge was expressed as protective efficacy (PE) and ranged from 13% in the case of ECOK1_3473 to 82% in the case of ECOK1_3385 (Table 1).

Most of the protective candidates are encoded in large genomic islands and have just a putative or hypothetical function assigned. Among the protective candidates identified, two seem to be associated with a type 2 secretion system (T2SS), including ECOK1_3374 which corresponds to subunit K, believed to be a minor subunit of a pseudopilus (30). The most protective candidate, ECOK1_3385, is a putative lipoprotein that shares significant similarity with an accessory colonization factor (AcfD) of *Vibrio cholerae*, also secreted by a T2SS. The role of this protein in *V. cholerae* is still unclear, but it appears to be required for efficient colonization of infant mouse small intestine (31).

Other protective candidates have been previously described. Hemolysin A (HlyA, ecp_3827) is a toxin of the repeats-in-toxin

Table 1. The most protective candidates selected in the mouse model of sepsis

Candidate	Annotation	Protective efficacy, %	*P* value	Solubility	Location	Source
ECOK1_3385	Lipoprotein, putative	82	<0.0001	Soluble	PAI $V_{IHE3034}$	This study
ecp_3827	Hemolysin A	76	<0.0001	Insoluble	PAI I_{536}, PAI II_{536}	(53)
c1275	Hypothetical protein	45	0.0002	Soluble	PAI-CFT073-*serX*	(54)
c5321	Hypothetical protein	33	<0.0001	Soluble	—	This study
ECOK1_3457	TonB-dependent siderophore receptor	25	0.029	Insoluble	—	This study
c0975	Hypothetical protein	24	0.1	Insoluble	Φ-CFT073-b0847	(54)
ECOK1_3374	gspK general secretion pathway protein K	20	0.0009	Soluble	PAI $V_{IHE3034}$	This study
ECOK1_0290	Bacterial Ig-like domain (group 1) protein	15	0.048	Soluble	—	This study
ECOK1_3473	Fimbrial protein	13	0.09	Soluble	—	This study

Candidates ECOK1_3385, ECOK1_3457, ECOK1_3374, ECOK1_0290, and ECOK1_3473 have been amplified from IHE3034; candidates c1275, c5321, and c0975 have been amplified from CFT073; candidate ecp_3827 has been amplified from 536. The infection was performed with the homologous challenge strains. *P* value <0.05, calculated by Fisher's exact test, indicates a significant difference in survival between vaccinated and control groups.

family with membrane-permeabilizing activity, secreted by the type I secretion system (32), and shown to be protective against *E. coli* infection (33). FitA (ECOK1_3457) participates in iron uptake and is up-regulated in iron-restricted media (34). Other candidates may be involved in bacterial adhesion. For example, ECOK1_0290 shares 94% similarity with EaeH (35), a protein identified in an ETEC strain which has some similarity with EHEC intimin (*eae*) (30%). Although the sequence similarity with invasin and intimin is low (up to 35%), ECOK1_0290 likely shares a similar immunoglobulin-like folding and could be involved in cell adhesion. Another candidate, ECOK1_3473, is a putative fimbrial subunit potentially related to a recently described ExPEC adhesin EA/I involved in the colonization of chicken's lung (36). Screening the public databases, we found that most of the genes coding for the nine protective antigens are also present in other pathogenic *E. coli* strains, including extraintestinal and intestinal isolates (Fig. S2), suggesting that a combination of these antigens may be able to provide protection against most pathogenic *E. coli*.

Protective Antigen Encoded by ECOK1_3385 Is Expressed but Not Secreted by Nonpathogenic Strains. We focused our attention on the antigen encoded by ECOK1_3385, the most protective antigen identified during the primary screening. Fig. 2 shows that active immunization with the antigen (Fig. 2 *A* and *C*) or passive immunization with sera raised against it (Fig. 2 *B* and *D*) provided nearly complete protection from bacteremia and from mortality. To explore the conservation of ECOK1_3385, the full-length nucleotide sequence of the *ECOK1_3385* gene was determined in 96 ExPEC isolates. Analysis of the predicted amino acid sequences of the 96 genes and of others available in databases revealed an identity ranging from 86% to 100%. In addition, the antigen appears to cluster into two main variants (Fig. S3). To evaluate the ability of this antigen to cross-protect against heterologous strains, mice were immunized with the antigen derived from IHE3034 and then challenged with four ExPEC strains with different level of identity. As shown in Table 2, protection was observed even against the most distant variant of antigen ECOK1_3385. The protective efficacy was variable but ranged from 43% to 78%, suggesting that this antigen has potential to be broadly protective.

Table 2. Cross-protection of ECOK1_3385 against heterologous ExPEC strains

Strain	Identity, %*	Protective efficacy, %	*P* value
B616	100	78	0.0002
536	86	60	0.0320
9855/93	87	50	0.0065
BK658	98	43	0.28

*Aminoacid identity with ECOK1_3385 from strain IHE3034.

To understand the potential coverage of all pathogenic *E. coli* by a vaccine containing the antigen encoded by *ECOK1_3385*, we amplified the gene in 147 *E. coli* isolates including various pathotypes of human and animal origin. We found that the gene is present in 83% of the strains tested. To further investigate the distribution of *ECOK1_3385*, we amplified the gene in a total of 573 isolates representing the entire span of the *E. coli* phylogeny (Fig. 3). We showed that ECOK1_3385 is widely distributed among all strains with an overall higher presence in intestinal

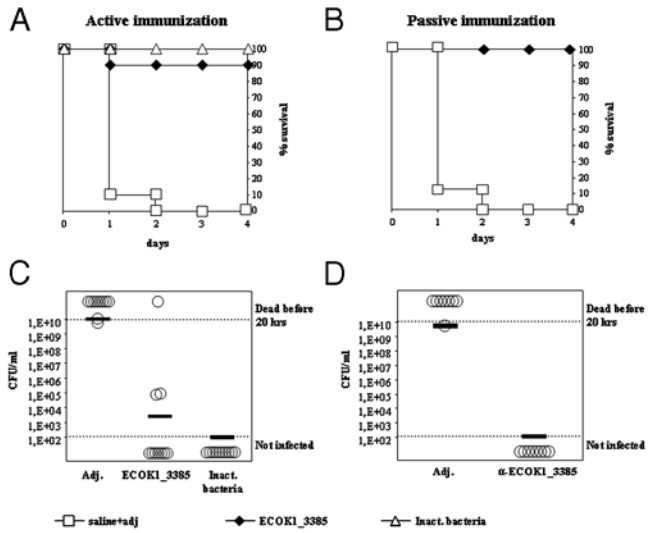

Fig. 2. Passive and active immunization using the antigen encoded by *ECOK1_3385*. Mice (10 per group) were actively (*A* and *C*) or passively (*B* and *D*) immunized with recombinant ECOK1_3385 (◆), heat-inactivated IHE3034 (△) and saline solution with Freund's adjuvant (□). Following challenge with the strain IHE3034 survival of mice was followed for 4 days (*A* and *B*). Bacterial counts in the blood (*C* and *D*) were determined at 20 h by counting the colony forming units (CFUs) in the blood detected by 10-fold serial dilutions and plating method. Bacteremia mean values in each group were determined assuming that mice, dead before 20 h, had the maximum level of CFU causing mouse death.

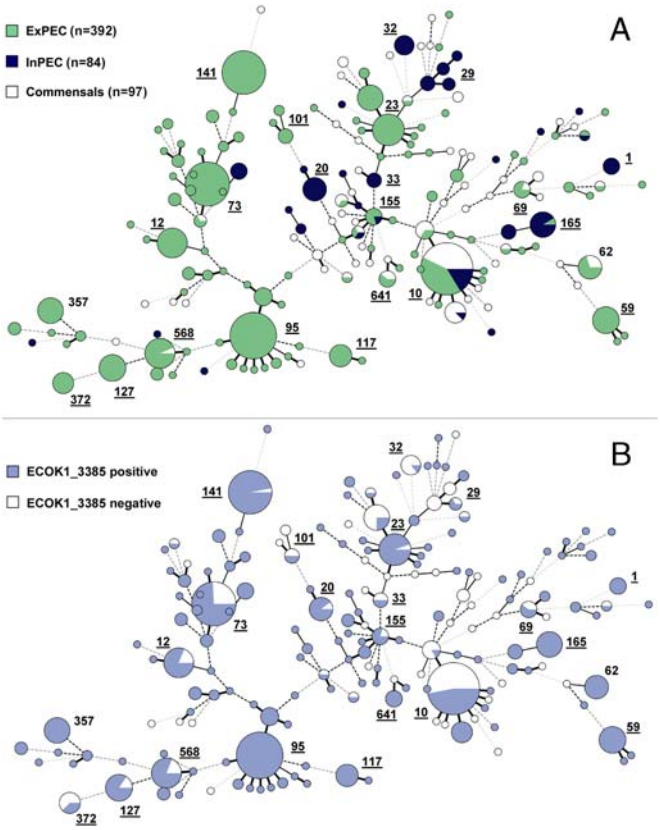

Fig. 3. MSTree analysis of ECOK1_3385 distribution in a collection of *E. coli* isolates. The MSTree was constructed by a graphical tool, implemented as part of Bionumerics V4.6 (Applied Maths Belgium), that links allele designations within an MLST database. ST complexes are defined as containing at least three STs, differing in no more than one allele to their nearest neighbor. The MSTree displays the quantitative relationships between STs (given as numbers) and ST complexes (given as underlined numbers), measured as the number of shared alleles, by lines of different thickness and type. (*A*) Distribution of different *E. coli* groups (ExPEC, extraintestinal pathogenic *E. coli*; InPEC, intestinal pathogenic *E. coli*; commensals) in the phylogenetic background of 573 strains. The groups have been classified not only on the basis of their clinical history but also according to extensive virulence typing. (*B*) Distribution of ECOK1_3385 in the phylogenetic background of 573 strains.

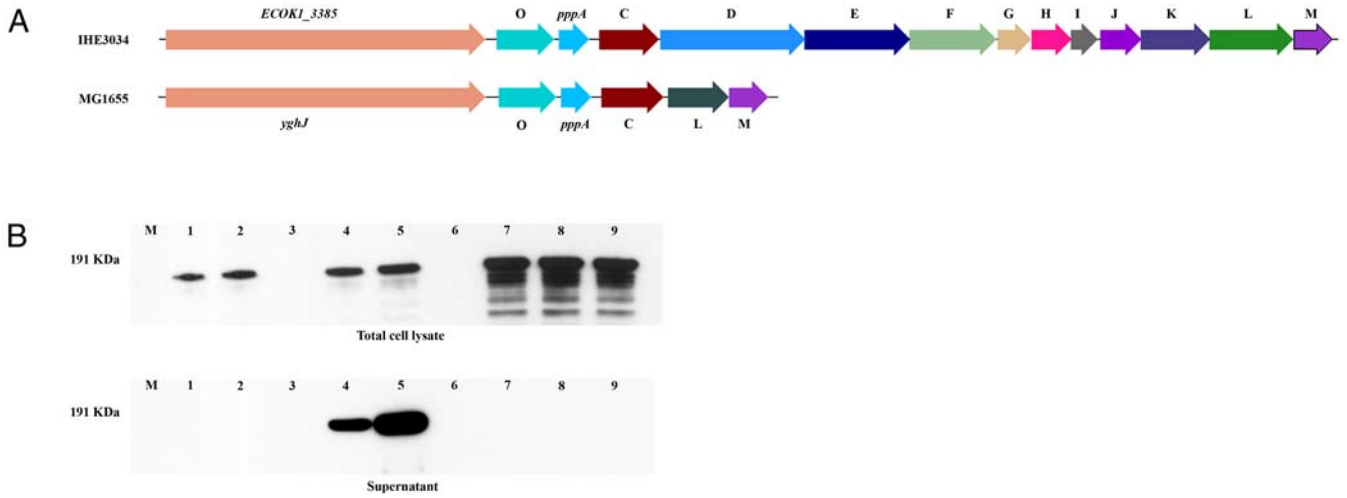

Fig. 4. T2SS is necessary for the secretion of the antigen encoded by *ECOK1_3385*. Map of the genomic region encoding ECOK1_3385 and the adjacent T2SS in strains IHE3034 and MG1655. The figure shows that the T2SS is truncated in the strain MG1655. Presence of the antigen in the total cell lysates and culture supernatants of strains, MG1655 (lane 1), W3110 (lane 2), CFT073 (lane 3), 536 (lane 4), IHE3034 (lane 5), IHE3034Δ*ECOK*1$_3$385 (lane 6), IHE3034Δ*ECOK*1$_3$381 (lane 7), IHE3034Δ*ECOK*1$_3$380 (lane 8), and IHE3034Δ*ECOK*1$_3$374 (lane 9). Antigen ECOK1_3385 is present in the total cell lysates of all strains with the exception of lanes 3 and 6 that contain strain CFT073 and the knockout mutant IHE3034Δ*ECOK*1$_3$385 used as negative controls. The antigen is secreted only in strains 536 and IHE3034 that have an intact T2SS (lanes 4 and 5, respectively). Deletion of T2SS components such as ECOK1_3381 (component D), ECOK1_3380 (component E), and ECOK1_3374 (component K) leads to the abrogation of its secretion but not expression (lanes 7, 8, and 9, respectively).

and extraintestinal pathogenic (70–83%) compared to commensal isolates (59%). However, it is important to consider that "commensals" are only poorly defined, and that the absence of *E. coli* pathogenic-specific genes does not imply that they are nonpathogenic.

A closer look at the region encoding ECOK1_3385 showed that the adjacent T2SS region is truncated in MG1655 (Fig. 4*A*), suggesting that the antigen encoded by *ECOK1_3385* may not be secreted in this strain. Western blot analysis, indeed, confirmed that in MG1655 and W3110 the antigen is expressed in total cell lysates, but not secreted (lanes 1 and 2 of Fig. 4*B*) and not surface associated, as demonstrated by Western blot analysis of outer membrane vesicles (OMV) fractions and by FACS analysis (Fig. S4). In the pathogenic strains 536 and IHE3034, the antigen is expressed and secreted into the supernatant (Fig. 4*B*, lanes 4 and 5). The secreted antigen ECOK1_3385 does not derive from OMV shedding because it is found in soluble form after ultracentrifugation. To confirm that a complete T2SS is necessary for the secretion of the antigen encoded by ECOK1_3385, three deletion mutants of the T2SS components were constructed in IHE3034. In particular, ECOK1_3381, the outer membrane secretin (component D), ECOK1_3380, a cytoplasmic ATPase that provides the energy required to regulate extracellular secretion (component E), and ECOK1_3374, a minor pilus subunit (component K) and also a protective candidate, were deleted. Deletion of each single component resulted in the abrogation of ECOK1_3385 secretion but not expression (Fig. 4*B*, lanes 7, 8, and 9, respectively), confirming that, indeed, the T2SS is necessary for the secretion of the antigen encoded by ECOK1_3385.

Discussion

The absence of a broadly protective vaccine against ExPEC strains is a major problem for modern society because some of the diseases caused by these bacteria, such as urinary tract infections (UTIs), are associated with high costs to health care systems and others, such as sepsis, are associated with high infant mortality (5). In addition, ExPEC infections are arguably increasing in susceptible subpopulations, such as patients with diabetes, multiple sclerosis, HIV/AIDS, catheters, and spinal cord injuries, pregnant women, children, and the elderly. The overall problem is exacerbated by the increasing antibiotic resistance and the number of recurrent infections (44%). The attempts to develop a broadly protective and safe vaccine against ExPEC have not been successful. The large antigenic and genetic variability of these strains, for example in the O antigen, has been a major obstacle to find common antigens. Adhesins associated with host interaction, for example P fimbriae (37) and type 1 pilus subunits (38), were able to confer protection in animal models but failed in phase II clinical trials. Whole-cell preparations are available against ExPEC (39) and killed genetic engineered strains are still a matter of study (40). Other antigens considered for vaccines are the outer membrane proteins associated with iron uptake, such as IroN (41), Hma, IreA, IutA (42), FyuA, and other hypothetical TonB-dependent receptors (43).

Because reverse vaccinology has shown to be a potential route to protective antigen identification where conventional approaches to vaccine development had failed (44, 45), we decided to search the whole genome for broadly protective antigens. To do so, we determined the genome sequence of a strain isolated from neonatal meningitis and we compared the sequence to the available genomes of ExPEC and nonpathogenic strains. We found that the genome of the K1 strain contained 19 genomic islands that were not present in the nonpathogenic strains. These different regions accounted for almost 20% of the total genome, indicating the huge diversity of *E. coli* strains. The most remarkable finding, however, was that these large genomic differences were not only present between the pathogenic and the nonpathogenic strains, but also within the pathogenic ExPEC strains, confirming the hypothesis of the individual virulence potential for each ExPEC strain (27).

In spite of the huge genomic diversity, by using a subtractive reverse vaccinology approach, we were able to identify nine potential vaccine candidates against ExPEC able to confer a protection efficacy ranging from 13% to 82% in a sepsis mouse model by active immunization and challenge with the homologous strain. In addition, the finding that passive protection tests with immune sera against the best protective antigen ECOK1_3385 show almost complete protection provides evidence that antibodies may be the key effectors for protection. Further studies will shed light on the type of protective immune response mediated by the vaccine candidates.

Most of the identified antigens have just a hypothetical function assigned. Some of them are conserved not only in ExPEC strains but also in pathogenic intestinal *E. coli,* suggesting that

they may be useful for a broadly cross-reactive *E. coli* vaccine. Antigens were selected mostly from the regions absent in nonpathogenic *E. coli* strains. However, in the case of ECOK1_3385, which is the most protective antigen, we found that the gene was present, reasonably well conserved, and expressed also in nonpathogenic organisms, although the antigen could not be secreted because the adjacent T2SS was not functional. The finding that some nonpathogenic strains produce but not secrete the antigen encoded by ECOK1_3385 suggests that this antigen may be associated with virulence.

In conclusion, by sequencing the genome of an ExPEC strain causing neonatal meningitis, and comparing the genome to those of other ExPEC and nonpathogenic strains, we have confirmed the great diversity of *E. coli* genomes and shown that the diversity is not only between nonpathogenic and pathogenic strains, but also within strains causing similar diseases. In spite of the observed diversity, we have been able to identify nine protective antigens that are conserved among several *E. coli* strains suggesting that, by combining some of these antigens, we may be able to develop a broadly cross-reactive vaccine against ExPEC. Preliminary results show that some of the antigens are also present in intestinal pathogenic *E. coli*, suggesting that the vaccine may have the potential to cover most pathogenic *E. coli*.

Materials and Methods

Strains. IHE3034 (O18:K1:H7), ST95, is a neonatal meningitis-associated strain isolated in Finland in 1976 (8). The IHE3034 strain sequenced in this work has lost its plasmids during laboratory passages and is more virulent than the ancestor strain in the sepsis mouse model.

Open Reading Frame (ORF) Prediction and Gene Identification. ORFs were predicted as previously described (44, 45). All ORFs were searched with FASTA3 against all ORFs from the complete genomes and matches with a fasta *P* value of 10^{-15} were considered significant.

Antigens Selection. Antigens were selected based on different criteria: low identity between ExPEC (IHE3034, 536, and CFT073) and nonpathogenic (DH10B, W3110, and MG1655) ORFs, low number of transmembrane domains (TMDs), and subcellular localization. Proteins sharing an identity and overlap with nonpathogenic strains above 90% and 80%, respectively, were discarded, as well as proteins with a cytoplasmic prediction and a high number of TMDs (>3). Protein subcellular localization was predicted by bioinformatic tools (46, 47) and also proteomic approaches (48).

Cloning and Expression of Candidates. Antigens were PCR amplified from the genomic DNA templates, cloned in pET-21b vector (Novagen) and transformed in DH5α-T1 chemically competent cells for propagation (Invitrogen). BL21(DE3) chemically competent cells were used for protein expression. All candidates were cloned and expressed as His-tagged fusion proteins without the predicted signal sequence, and the proteins purified as previously described (44).

Mice Immunization and Lethal Challenge. Active protection was evaluated in a sepsis mouse model. CD1 out-bred mice (8–10 females per group, at least three groups per antigen, 5 weeks old, purchased from Charles River Italia) were immunized by subcutaneous injections at day 1, 21, and 35 with 20 μg of recombinant protein in 150 μL of saline solution using complete and incomplete Freund's (Sigma) as adjuvant. Positive control mice were immunized with 10^8 heat-inactivated bacteria (65 °C for 30 min) in Freund's adjuvant, whereas negative control mice were immunized with Freund's adjuvant in saline solution. Immunized animals were challenged at day 49 with a lethal dose of homologous strains IHE3034, CFT073, or 536, causing 80% to 100% killing in control mice. Heparinised-blood samples were collected from survived mice at 20 h after challenge to determine bacteremia levels and the mortality was monitored for 4 days after challenge.

For passive protection experiments, mice were immunized intravenously with 100 μL of rabbit hyperimmune sera and infected intraperitoneally 24 h after immunization as described above.

Vaccine protection was expressed as percent PE calculated as (1 − [% dead in vaccine group/% dead in negative control group] × 100). Statistical analysis was carried out by two-tailed Fisher's exact test comparing the number of survived animals in vaccinated groups with the number of survived animals in negative control groups. *P* value <0.05 are considered as significant.

PCR Amplification and Sequencing of ECOK1_3385. Genomic DNA was prepared by culturing bacteria overnight at 37 °C in atmosphere humidified with 5% CO_2 in LB (Difco). Chromosomal DNA was prepared from 1.5 mL of culture using the GenElute Bacterial Genomic DNA Kit (Sigma) according to the manufacturer's instructions. DNA concentration was calculated by optical density determination at 260 nm. About 100 ng of chromosomal DNA was used as template for the amplification of *ECOK1_3385* gene. The amplification enzyme used was the Phusion® DNA Polymerase (Finnzymes). All genes were amplified using primers external to the coding region. Primers were designed in conserved DNA region and the sequences are reported in Table S2. *ECOK1_3385* was amplified using primers ECOK1_3385_1 and ECOK1_3385_22. PCR conditions were as follows: 35 cycles of denaturation at 98 °C for 10 s, annealing at 55 °C for 20 s, and elongation at 72 °C for 3 min. PCR products were purified with Agencourt® AMPure® protocol (Beckman Coulter) and sequenced on the capillary sequencer ABI3730xl DNA Analyzers (Applied Biosystems). Sequences were assembled with Sequencher 4.8 (Gene Codes) and aligned and analyzed using the Vector NTI Suite 10.

Multilocus Sequence Typing (MLST) Phylogenetic Analysis. MLST analyses were performed according to a previously published protocol (49) using oligonucleotide primers given on the MLST website (http://mlst.ucc.ie/mlst/mlst/dbs/Ecoli/). MLST data of *E. coli* strains were partly adopted from previous studies (50, 51).

Bioinformatic Tools. Comparative genome analysis were performed using the Genetics Computer Group version 11.1 (Accelrys, Inc.). Circular maps were plotted using GenomeViz (52). Amino acid alignments were performed using CLUSTALW and the phylogenetic tree was inferred from the alignments by the neighbor-joining distance-based method and bootstrapped 100 times.

ACKNOWLEDGMENTS. We thank Mary Kim, Hue Vuong, Jessica Hostetler for gap closure and Sean C. Daugherty, William C. Nelson, Lauren M. Brinkac, Robert J. Dodson, Scott A. Durkin, Ramana Madupu, Derek M. Harkins, and Susmita Shrivastava for automated and manual annotation. We are grateful to Marina Cerquetti (Istituto Superiore di Sanitá) for kindly providing ExPEC clinical isolates and to Ulrich Vogel and Karin Schnetz for providing *E. coli* isolates. We thank Giacomo Matteucci, Marco Tortoli, and the animal care facility for technical assistance in the experimentation. We thank Giorgio Corsi for artwork and Catherine Mallia for manuscript editing. We kindly thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior for D.G.M. PhD fellowship (2111-03-9). G.D. and D.P. were supported by the Wellcome Trust. L.H.W. was funded by Bundesministerium fur Bildung und Forschung, Project Food-Borne Zoonotic Infections of Humans (FBI-Zoo) (Grant 01KI07120).

1. Kaper JB, Nataro JP, Mobley HL (2004) Pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2 (2):123–140.
2. Blattner FR, et al. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277(5331):1453–1474.
3. Ogura Y, et al. (2009) Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. *Proc Natl Acad Sci USA* 106(42):17939–17944.
4. Ahmed N, Dobrindt U, Hacker J, Hasnain SE (2008) Genomic fluidity and pathogenic bacteria: applications in diagnostics, epidemiology and intervention. *Nat Rev Microbiol* 6(5):387–394.
5. Russo TA, Johnson JR (2003) Medical and economic impact of extraintestinal infections due to *Escherichia coli*: Focus on an increasingly important endemic problem. *Microbes Infect* 5(5):449–456.
6. Russo TA, Johnson JR (2006) Extraintestinal isolates of *Escherichia coli*: Identification and prospects for vaccine development. *Expert Rev Vaccines* 5(1):45–54.
7. Moriel DG, et al. (2008) Genome-based vaccine development—A shortcut for the future. *Hum Vaccines* 4(3):184–188.
8. Achtman M, et al. (1983) Six widespread bacterial clones among *Escherichia coli* K1 isolates. *Infect Immun* 39(1):315–335.
9. Bingle LE, Bailey CM, Pallen MJ (2008) Type VI secretion: A beginner's guide. *Curr Opin Microbiol* 11(1):3–8.
10. Dobrindt U, et al. (2001) S-Fimbria-encoding determinant *sfa*(I) is located on pathogenicity island III(536) of uropathogenic *Escherichia coli* strain 536. *Infect Immun* 69(7):4248–4256.
11. Ulett GC, et al. (2007) Functional analysis of antigen 43 in uropathogenic *Escherichia coli* reveals a role in long-term persistence in the urinary tract. *Infect Immun* 75 (7):3233–3244.
12. Russo TA, et al. (2002) IroN functions as a siderophore receptor and is a urovirulence factor in an extraintestinal pathogenic isolate of *Escherichia coli*. *Infect Immun* 70 (12):7156–7160.

13. Sabri M, et al. (2008) Contribution of the SitABCD, MntH, and FeoB metal transporters to the virulence of an avian pathogenic *Escherichia coli* (APEC) O78 strain. *Infect Immun* 76(2):601–611.
14. Schubert S, Picard B, Gouriou S, Heesemann J, Denamur E (2002) Yersinia high-pathogenicity island contributes to virulence in *Escherichia coli* causing extraintestinal infections. *Infect Immun* 70(9):5335–5337.
15. Heywood W, Henderson B, Nair SP (2005) Cytolethal distending toxin: Creating a gap in the cell cycle. *J Med Microbiol* 54(Pt 3):207–216.
16. Nougayrede JP, et al. (2006) *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* 313(5788):848–851.
17. Sandt CH, Hopper JE, Hill CW (2002) Activation of prophage *eib* genes for immunoglobulin-binding proteins by genes from the IbrAB genetic island of *Escherichia coli* ECOR-9. *J Bacteriol* 184(13):3640–3648.
18. Wieler LH, et al. (1996) The enterohemolysin phenotype of bovine Shiga-like toxin-producing *Escherichia coli* (SLTEC) is encoded by the EHEC-hemolysin gene. *Vet Microbiol* 52(1–2):153–164.
19. Cianciotto NP (2005) Type II secretion: A protein secretion system for all seasons. *Trends Microbiol* 13(12):581–588.
20. Pluschke G, Achtman M (1985) Antibodies to O-antigen of lipopolysaccharide are protective against neonatal infection with *Escherichia coli* K1. *Infect Immun* 49 (2):365–370.
21. Zou Y, He L, Chi F, Jong A, Huang SH (2008) Involvement of *Escherichia coli* K1 *ibeT* in bacterial adhesion that is associated with the entry into human brain microvascular endothelial cells. *Med Microbiol Immunol* 197(4):337–344.
22. Huang SH, Wan ZS, Chen YH, Jong AY, Kim KS (2001) Further characterization of *Escherichia coli* brain microvascular endothelial cell invasion gene *ibeA* by deletion, complementation, and protein expression. *J Infect Dis* 183(7):1071–1078.
23. Johnson TJ, et al. (2007) The genome sequence of avian pathogenic *Escherichia coli* strain O1:K1:H7 shares strong similarities with human extraintestinal pathogenic *E. coli* genomes. *J Bacteriol* 189(8):3228–3236.
24. Touchon M, et al. (2009) Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* 5(1):e1000344.
25. Chen SL, et al. (2006) Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: A comparative genomics approach. *Proc Natl Acad Sci USA* 103(15):5977–5982.
26. Welch RA, et al. (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci USA* 99 (26):17020–17024.
27. Brzuszkiewicz E, et al. (2006) How to become a uropathogen: Comparative genomic analysis of extraintestinal pathogenic *Escherichia coli* strains. *Proc Natl Acad Sci USA* 103(34):12879–12884.
28. Mokady D, Gophna U, Ron EZ (2005) Extensive gene diversity in septicemic *Escherichia coli* strains. *J Clin Microbiol* 43(1):66–73.
29. Bielaszewska M, et al. (2007) Aspects of genome plasticity in pathogenic *Escherichia coli*. *Int J Med Microbiol* 297(7–8):625–639.
30. Sandkvist M (2001) Biology of type II secretion. *Mol Microbiol* 40(2):271–283.
31. Peterson KM, Mekalanos JJ (1988) Characterization of the *Vibrio cholerae* ToxR regulon: identification of novel genes involved in intestinal colonization. *Infect Immun* 56(11):2822–2829.
32. Gentschev I, Dietrich G, Goebel W (2002) The *E. coli* alpha-hemolysin secretion system and its use in vaccine development. *Trends Microbiol* 10(1):39–45.
33. O'Hanley P, Lalonde G, Ji G (1991) Alpha-hemolysin contributes to the pathogenicity of piliated digalactoside-binding *Escherichia coli* in the kidney: Efficacy of an alpha-hemolysin vaccine in preventing renal injury in the BALB/c mouse model of pyelonephritis. *Infect Immun* 59(3):1153–1161.
34. Ouyang Z, Isaacson R (2006) Identification and characterization of a novel ABC iron transport system, *fit*, in *Escherichia coli*. *Infect Immun* 74(12):6949–6956.
35. Chen Q, Savarino SJ, Venkatesan MM (2006) Subtractive hybridization and optical mapping of the enterotoxigenic *Escherichia coli* H10407 chromosome: Isolation of unique sequences and demonstration of significant similarity to the chromosome of *E. coli* K-12. *Microbiology* 152(Pt 4):1041–1054.
36. Antao EM, et al. (2009) Signature-tagged mutagenesis in a chicken infection model leads to the identification of a novel avian pathogenic *Escherichia coli* fimbrial adhesin. *PLoS One* 4(11):e7796.
37. Roberts JA, et al. (2004) Antibody responses and protection from pyelonephritis following vaccination with purified *Escherichia coli* PapDG protein. *J Urol* 171 (4):1682–1685.
38. Langermann S, et al. (1997) Prevention of mucosal *Escherichia coli* infection by FimH-adhesin-based systemic vaccination. *Science* 276(5312):607–611.
39. Russo TA, Carlino-Macdonald U (2008) Extraintestinal pathogenic isolates of *Escherichia coli* do not possess active IgA(1), IgA(2), sIgA or IgG proteases. *FEMS Immunol Med Microbiol* 53(1):65–71.
40. Russo TA, et al. (2007) A killed, genetically engineered derivative of a wild-type extraintestinal pathogenic *E. coli* strain is a vaccine candidate. *Vaccine* 25(19):3859–3870.
41. Russo TA, et al. (2003) The Siderophore receptor IroN of extraintestinal pathogenic *Escherichia coli* is a potential vaccine candidate. *Infect Immun* 71(12):7164–7169.
42. Alteri CJ, Hagan EC, Sivick KE, Smith SN, Mobley HL (2009) Mucosal immunization with iron receptor antigens protects against urinary tract infection. *PLoS Pathog* 5(9): e1000586.
43. Durant L, et al. (2007) Identification of candidates for a subunit vaccine against extraintestinal pathogenic *Escherichia coli*. *Infect Immun* 75(4):1916–1925.
44. Pizza M, et al. (2000) Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 287(5459):1816–1820.
45. Tettelin H, et al. (2002) Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. *Proc Natl Acad Sci USA* 99(19):12391–12396.
46. Nakai K, Kanehisa M (1991) Expert system for predicting protein localization sites in gram-negative bacteria. *Proteins* 11(2):95–110.
47. Gardy JL, et al. (2005) PSORTb v.2.0: Expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics* 21(5):617–623.
48. Scorza FB, et al. (2008) Proteomics characterization of outer membrane vesicles from the extraintestinal pathogenic *Escherichia coli* Δ*tolR* IHE3034 mutant. *Mol Cell Proteomics* 7(3):473–485.
49. Wirth T, et al. (2006) Sex and virulence in *Escherichia coli*: An evolutionary perspective. *Mol Microbiol* 60(5):1136–1151.
50. Sankar TS, et al. (2009) Fate of the H-NS-repressed *bgl* operon in evolution of *Escherichia coli*. *PLoS Genet* 5(3):e1000405.
51. Ewers C, Antao EM, Diehl I, Philipp HC, Wieler LH (2009) Intestine and environment of the chicken as reservoirs for extraintestinal pathogenic *Escherichia coli* strains with zoonotic potential. *Appl Environ Microbiol* 75(1):184–192.
52. Ghai R, Hain T, Chakraborty T (2004) GenomeViz: Visualizing microbial genomes. *BMC Bioinformatics* 5:198.
53. Dobrindt U, et al. (2002) Genetic structure and distribution of four pathogenicity islands (PAI I(536) to PAI IV(536)) of uropathogenic *Escherichia coli* strain 536. *Infect Immun* 70(11):6365–6372.
54. Lloyd AL, Rasko DA, Mobley HL (2007) Defining genomic islands and uropathogen-specific genes in uropathogenic *Escherichia coli*. *J Bacteriol* 189(9):3532–3546.