UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Transport proteins promoting E. coli pathogenesis /

Permalink

https://escholarship.org/uc/item/05d6c0bq

Author

Tang, Fengyi

Publication Date

2013

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Transport proteins promoting *E. coli* pathogenesis

A Thesis submitted in partial satisfaction of the requirements for the degree Master of Science

in

Biology

by

Fengyi Tang

Committee in charge:

Professor Milton Saier Jr., Chair
Professor Eric Allen
Professor Julian Schroeder

2013

Copyright

Fengyi Tang, 2013

All rights reserved.

The Thesis of Fengyi Tang is approved and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____
Chair

University of California, San Diego

2013

TABLE OF CONTENTS

Signature Page……..... iii

Table of Contents……..... iv

List of Figures……..... v

List of Tables……..... vi

Acknowledgements……..... vii

Abstract of the Thesis……..... ix

Introduction……..... 1

Methods……..... 4

Results……..... 8

Discussion……..... 39

References……..... 51

Appendix……..... 64

LIST OF FIGURES

Figure 1. Distribution of transport proteins according to TC class (left) & TC subclass (right) for *E. coli* K12 str. MG1655 (top) and the seven pathogenic strains combined (bottom).
…….. 75

Figure 2. Total number of proteins associated with iron transport in the eight *E. coli* strains in which (A) shows the relationship between the serovars (x-axis) and number of iron-siderophore transporters (y-axis).
…….. 78

Figure 3. Distribution of transport proteins organized according to TC class (left) & TC subclass (right) for each strain examined: (A) *E. coli* 559, (B) ABU, (C) APE, (D) CFT, (E) E24, (F) UMN, (G) O15, (H) K12
……..……..... 80

LIST OF TABLES

Table 1. Seven *E. coli* pathovars, the non-virulent *E. coli* K12 *s*train, and their basic traits……..……..... 64

Table 2. Numbers of proteins (even numbered columns) and percent distribution of subclasses by TC number (odd numbered columns) for all eight strains of *E. coli* as well as numbers of transport proteins per strain (bottom) and all strains combined.....……..... 65

Table 3. Toxins identified in *E. coli* pathovars and K12……..... 66

Table 4. Pathogenicity-related transenvelope protein secretion systems……..... 67

Table 5. Outer membrane protein secretion systems……..... 68

Table 6. Outer membrane iron-siderophore receptors……..... 69

Table 7. Inner membrane iron uptake systems ……... 70

Table 8. Number of OMR and ABC transport proteins associated with iron transport... 71

Table 9. Metabolite transporters unique to intracellular vs. extracellular pathogens….. 72

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Milton Saier Jr., Ph.D. for his mentorship and support during my research and education. Watching him embody the heart and soul of this lab has taught me what self-actualization looks like through the lenses of an empowered individual. I would also like to acknowledge the distinguished individuals of Saier lab whose presence and support have motivated me to apply myself each day. These include: Andre, Andrew K, Scott M, Ida, Andrew L, Bora, Dan, Bruce, Ross C, and Ake. I would like to acknowledge Vamsee Reddy for writing a program for tabulating the results summarized in supplementary table S2. This work was supported by NIH grant GM077402. I would also like to thank the only child policy in China implemented in the 60's or 70's, I don't quite recall. This policy gave me substantial pressure to push myself to the edge of my abilities and to utilize each minute of the day wisely, knowing well that I may be the sole breadwinner of my family one day. I would also like to acknowledge the lyrics of "Don't You Worry Child," produced by the *Swedish House Mafia (Ake Westernfield)*, which kept me going through tough times. The lyrics are listed below:

There was a time,

When I used to look into my father's eyes.

In a happy home,

I was the king I had a golden throne.

But those days are gone,

with memories painted on the walls.

I could still hear the sound,

of the places where I was born.

Up a hill across a blue lake,

where I had my first heartbreak.

I still remembered how it all changed.

My father said,

"Don't you worry, don't you worry child,

See Heaven's got a plan for you.

Don't you worry, don't you worry now."

I would also like to acknowledge GNC's Optimum Nutrition Gold Standard 100% Whey Protein, whose nutrients kept me swoll, allowing me to live up to the expectations of "brain and brawn."

Last, but not least, I am extremely grateful for my parents' unreserved support. Without them, my education would not be possible.

ABSTRACT OF THE THESIS

Transport proteins promoting *E. coli* pathogenesis.

by

Fengyi Tang

Master of Science in Biology

University of California, San Diego, 2013

Professor Milton Saier Jr., Chair

Escherichia coli is a genetically diverse species infecting hundreds of millions of people worldwide annually. We examined seven well-characterized *E. coli* pathogens causing urinary tract infections, gastroenteritis, pyelonephritis and haemorrhagic colitis. Their transport proteins were identified and compared with each other and a non-pathogenic *E. coli* K12 strain to identify transport proteins related to pathogenesis. Each pathogen possesses a unique set of protein secretion systems for export to the cell surface or for injecting effector proteins into host cells. Pathogens have increased numbers of iron siderophore receptors and ABC iron uptake transporters, but the numbers and types

of low-affinity secondary iron carriers were uniform in all strains. The presence of outer membrane iron complex receptors and high-affinity ABC iron uptake systems correlated, suggesting co-evolution. Further, increased numbers of TonB and TolA receptor energizers were observed in those pathogens having increased iron acquisition capabilities. Each pathovar encodes a different set of pore-forming toxins and virulence-related outer membrane proteins lacking in K12. Intracellular pathogens proved to have a characteristically distinctive set of nutrient uptake porters, different from those of extracellular pathogens. The results presented in this report provide information about transport systems relevant to various types of *E. coli* pathogenesis that can be exploited in future basic and applied studies.

INTRODUCTION

E. coli has been and continues to be a major source of infant mortality, adult diarrhoea and urinary tract infections worldwide [1] [2]. Uropathogenic strains such as *E. coli* CFT073 are responsible for up to 80% of urinary tract infections (UTIs) and a majority of pyelonephritis cases [3]. Enteroaggregative and enterotoxic pathovars have been the leading causes of travellers' diarrhoea, which has significantly contributed to infant mortality [4]. Enterohaemorrhagic serotypes are currently the leading cause of haemorrhagic colitis which can lead to fatalities [5]. Neonatal meningitis strains approach 40% fatality rates in infected newborns and can cause permanent neurological problems even after successful treatment [6].

Some probiotic strains of *E. coli* contribute beneficially to human and other mammalian physiology. Prevalent examples include the *E. coli* Nissle and O83 strains, which have demonstrably alleviated colitis in mice [7]. Furthermore, studies have shown that Nissle 1917 reverses effects of enteroaggregative strains of *E. coli* such as LF82 [8]. Current research on probiotic *E. coli* in humans has proven to be clinically useful in the treatment of inflammatory bowel disease (IBD) syndromes [9, 10]. These strains reveal the tremendously varied effects *E. coli* can have on humans and other animals, a consequence of its genetic diversity.

Eight prominent types of *E. coli* pathogens have been identified based on varying mechanisms of infection. There are five types of diarrhoeagenic strains that pertain

specifically to the alimentary system: enteropathogenic (EPEC), enterohaemorrhagic (EHEC), enterotoxic (ETEC), enteroinvasive (EIEC), and enteroaggregative (EAEC). Diffusely adherent (DAEC) strains can infect both the urinary tract and the gastrointestinal tract due in part to their diverse mechanisms of host adhesion [6]. Extraintestinal *E. coli* pathovars (ExPEC) can be classified as uropathogenic (UPEC) or neonatal meningitis-causing (NMEC) [6]. In our study, the following representative pathogenic *E. coli* strains were examined: 55989 (EAEC), ABU 83972 (UPEC), APEC01 (ExPEC), CFT073 (UPEC), E24377A (ETEC), UMN026 (UPEC), and O157:H7 (EHEC). Nonpathogenic *E. coli* K12 str. MG1655 was included for comparative purposes to identify transport proteins and mechanisms likely to be unrelated to pathogenicity. In this communication, we use the first three letters/numbers of these strain designations as abbreviations to refer to these strains. Table 1 presents these abbreviations as well as basic information about the eight strains of *E. coli* included in our study.

The general paradigm of the pathogenic mechanism associated with *E. coli* virulence involves (1) adhesion, (2) protein injection into host cells, (3) subversion of signaling mechanisms and (4) colonization leading to impaired immune responses, membrane potential disruption and cytoskeletal manipulation [11]. All pathogens except EIEC utilize some type of adhesive protein to attach to host cells. EIEC utilizes host immuno delivery to escape to the basolateral side of the submucosa [12]. During the injection stage, *E coli* typically secretes effectors into host cells, using one of several protein secretion systems, either to escape the immune response or to alter signaling pathways. Host signaling interference usually involves hijacking normal host processes.

Typically, changes in the cytoskeleton are observed as *E. coli* sustains adhesion to the host mucosal surface or facilitates endocytosis of the bacterium for infiltration. Host membrane potential disruption and apoptosis induction are employed by several pathovars (EHEC, EPEC, ETEC). Although the types of adhesins, protein secretion systems, and effectors differ among the various pathovars, the general virulence paradigm is followed.

Virulence genes are often clustered in "pathogenicity islands" that reflect horizontal transfer across various lineages of *E. coli* as well as other bacteria [13]. Horizontal gene transfer has allowed *E. coli* to survive under a variety of *in vivo* conditions. It is hypothesized that selective pressures encourage more virulent strains to emerge, increasing the virulence of existing pathogens and creating new ones [14]; [15]. Because the pan genome of *E. coli* far exceeds the core genome, perhaps by as much as 20-fold [16], this increase in genetic plasticity allows for fluctuations in gene content, further contributing to divergence of *E. coli* pathogenic mechanisms [17].

METHODS

The proteomes of eight *E. coli* strains were screened for homologues of all proteins contained in the Transporter Classification Database (TCDB; www.tcdb.org) as of September, 2012 using a program designed for this purpose, G-BLAST [18]. The program retrieves information for both the genome query and the TC top hit sequences, TC#, protein size in number of amino acyl residues, number of predicted TMSs using the HMMTop 2.0 Program, e-value for the query and hit proteins, regions of sequence similarity and regions of TMS overlap. FASTA-formatted protein sequences from the completed genome were used. Each putative open-reading frame was used as a query in the BLASTP software to search for homologous proteins in TCDB. The low complexity filter was not used as it is normally of value only for larger datasets including proteins with multiple repeat elements. In addition, each open reading frame [19] was scanned with the HMMTOP 2.0 program [20] to predict the number of putative transmembrane segments (TMSs). The Web-based Hydropathy, Amphipathicity and Topology (WHAT) program [21] was used with a window size of 19 residues and an angle of 100 degrees to display the hydropathy plot for individual proteins in order to resolve the differences in the numbers of TMSs between the proteins retrieved and their TCDB homologues. The plot generated by WHAT allows the user to judge if a program such as HMMTOP has missed a TMS or has predicted a TMS inappropriately. A cut-off value of 0.0001 was used with the G-BLAST program so proteins retrieved with larger values (greater sequence divergence) were not recorded. Proteins with no predicted TMSs were eliminated so that only integral membrane proteins, primarily multispanning membrane

proteins, were retrieved. Proteins with only an N-terminal signal sequence are numerous because these proteins include almost all periplasmic, outer membrane and secreted proteins that are exported via the general secretory pathway (Sec) or twin arginine translocase (TAT). The topological prediction programs often miss these TMSs, recording them to have zero TMSs. Consequently, the numbers retrieved were not reliable and were therefore not always recorded. For example, single TMS proteins such as extracytoplasmic solute binding receptors of ABC transport systems were often predicted to lack a TMS, and therefore these proteins were not included in our study of the integral membrane transport proteins.

Proteins retrieved between the values of 0.0001 and 0.1 were examined manually to determine the likelihood that these proteins were members of recognized transport protein families, or if they might comprise representatives of novel families of putative transport proteins. A total of 82 non-orthologous homolog proteins were retrieved using the 0.0001 to 0.1 cutoff, but only 10 proved to be recognizable transport proteins. These were incorporated into TCDB. The 10 proteins were manually examined by conducting searches as follows. (1) TC-BLAST searches provided preliminary evidence for family assignment. (2) NCBI BLAST searches provided confirmation or refutation of family assignment based on the conserved domain database (CDD) and hits obtained with values to the query sequence of less than 1×10^{-7}. (3) Topological analyses revealed similarities and differences between the query sequence and members of the assigned family. (4) Proteins proving to represent new potential families were included in TC subclass 9.B.

Candidate proteins were subsequently examined in greater detail to estimate their substrate specificities. On the basis of the numbers and locations of TMSs as well as degrees of sequence similarity with entries of known function in TCDB, transport proteins were classified into families and subfamilies of homologous transporters according to the classification system presented in TCDB. Regions of sequence similarity were examined using the WHAT program which shows hydropathy plots to ensure that homology was in a transmembrane region of 3 or more TMSs and not only in hydrophilic domains. Proteins encoded within single multi-cistronic operons were often identified in order to gain evidence for multicomponent systems and to help deduce probable functions. Operon analyses (genome context, a.k.a., synteny analyses) were performed for candidate proteins with assigned or unassigned transport functions as described in Castillo and Saier (2010) and Reddy et al. (2012) [22] and [23].

Transport proteins thus obtained were systematically analyzed for unusual properties using published [18] and unpublished in-house software. Among the programs described by Reddy and Saier [18], used in this report, were the GSAT, Protocol1, Protocol2, TSSearch, SSearch and GBlast programs. Unpublished software was used to tabulate information according to TC# or other criteria such as substrate type. Unusual characteristics of the query *E. coli* sequences were identified based in part on topologies that differed from corresponding family members in TCDB as well as e-values obtained with G-BLAST. Unusual properties can result from events such as genetic deletion and fusion, sometimes resulting in the gain or loss of extra domains or the generation of

multifunctional proteins. Such results can be reflective of the actual protein sequence, but they can also be artifactual, due to sequencing errors or incorrect initiation codon assignment. In the latter cases, but not the former, the protein sequences were either corrected when possible or eliminated from our study.

Orthologous relationships among homologues were estimated by measuring the percent identities. We initially used an arbitrary cutoff of 95% identity, a value appropriate in view of the similarities of the strains considered. Homologues with lower percent identities were examined manually, revealing occasional sets of orthologs that exhibited less than 95% identity.

RESULTS

Eight *E. coli* genomes were analyzed for the occurrence of transport proteins using the Transporter Classification Database (TCDB; www.tcdb.org, [24]) and the Gblast program [18]. The results are summarized according to TC subclass in Table 2. Examining the total number of transport proteins present in these 8 genomes, we see that *E. coli* K12 has the fewest at 1128. The most found in any genome is 1287, 159 more than in K12. However, the seven pathogenic species combined contained 718 unique transport proteins lacking in K12.

TC subclass 1.A in TCDB includes all α-type channels. K12 has 39 such channel proteins, and the largest number found in any of the genomes was 42 for 559 and E24. Only CFT has less than K12 with 37. Surprisingly, however, the pathogenic organisms possess 24 non-orthologous channel proteins not found in K12 (Table 2). From these observations, it appears unlikely that α-type channels of subclass 1A are major contributors to *E. coli* pathogenesis.

TC subclass 1.B includes outer membrane β-type porins. In this subclass, we find striking differences between the different organisms, where K12 has far fewer than any of the other strains. The largest number of these porins is 108 observed in ABU. Therefore, it is not surprising that the seven pathogens possess 151 porin proteins non-orthologous to those in K12 (Table 2). This observation suggests that many outer membrane porins contribute to pathogenesis.

TC subclass 1.C includes pore-forming toxins. K12 has only nine such proteins, and E24 has just two more at 11. However, all the other organisms have between 15-20 such toxins, and 51 toxins non-orthologous to those in K12 were identified (Table 2). As expected, an array of toxins characterizes each of the pathogens. It should be noted that toxins that do not form transmembrane pores were not included in our study.

Holins comprise TC subclass 1.E. These small α-type channel-formers function in release of auto-lysins which allows phage particle release when phage-encoded, or self-destruction by a suicide-type process when chromosomally encoded [18]. K12 has only four such holins, and E24 has just three. The maximum found in any genome is seven, as observed for UMN. However, 19 of these proteins are non-orthologous to those found in K12 (Table 2). It is possible therefore that chromosomally encoded holins play important roles in pathogenicity, possibly in programmed cell death as an aid to biofilm formation and communal life ([25]; [26]; [27]).

The largest number of transport systems, but not the largest number of transport proteins, includes secondary carriers belonging to TC subclass 2.A. The eight strains have anywhere from 309 to 325 such systems. K12 has 311, and 96 proteins non-orthologous to those in K12 were identified. As will be shown below, many of these are likely to contribute to specific types of pathogenicity.

TC subclass 2.C comprises energizers for motility and outer-membrane transport. All strains have anywhere from 3-5 such proteins. E24 has only three while K12 and several of the other organisms have four. Nevertheless, the pathogens apparently possess six non-orthologous proteins not identified in K12. These include homologs of TonB (TC#2.C.1.1), TolA (TC#2.C.1.2.1), and MotAB (TC#1.B.30.1.1) systems and are likely to be important for virulence.

TC subclass 3.A are pyrophosphate-hydrolysis driven primary active transporters, usually multi-component systems. This is the category that has the largest number of proteins, although as noted above, not the largest number of transport systems. *E. coli* K12 has the fewest such systems (340), while UMN has the most (390), and 155 proteins found in pathogens appear to have no orthologs in K12. As will be described below, many of these systems play essential roles in virulence.

TC subclass 3.B includes decarboxylation driven sodium pumps. The strains examined have 3-5 such systems. As indicated in Table 2, K12 has four such proteins, and all the strains combined have just one more non-orthologous member. Thus, these Na^+-pumping decarboxylases probably play a housekeeping function in the maintenance of ionic homeostasis and are not a characteristic exclusive to the pathogens.

As expected, subclass 3.D, ion (H^+ or Na^+)-pumping electron carriers, are numerous in all organisms examined. However, there is substantial variation with APE having the least such proteins (59) while UMN has the most (72). K12 has 69, and only 7

electron carriers proved to be non-orthologous to those of K12. Thus, *E. coli* pathogens do not in general exhibit increased number of cation-pumping electron transport systems.

TC class 4 includes group translocators that are believed to modify their substrates in a process coupled to transport. Subclass 4.A includes the proteins of the multi-component sugar transporting phosphotransferase system (PTS) [28]. While K12 has the fewest such systems (49), both CFT and ABU have 13 more (62). In fact, 43 PTS proteins proved to be non-orthologous to those found in K12. These systems, which are likely to be important virulence factors, will be analyzed in greater detail below.

The 4.B subclass includes nicotinamide ribonucleoside (NR) group translocating uptake permeases (PnuC) although many homologs are believed to function by secondary active transport, acting on a large range of related compounds (see TCDB). The eight strains have between 2-4 such proteins with K12 having just two. Three such proteins appear to be non-homologous to those found in K12.

Subclass 4.C, fatty-acyl-coenzyme A ligases that activate fatty acids for lipid biosynthesis and may function in transport via group translocation [29], vary over two-fold in numbers. O15 has only seven, K12 has eight, and E24 has nine. However, ABU has 16 and CFT has 17. Surprisingly, only nine orthologs found in the seven pathogenic strains are lacking in K12. This means that all of the proteins lacking in K12 but present

in any one of the pathogens are present in CFT. It is possible that some of these systems play a role in virulence in some strains but not others.

A very different pattern is observed for the putative group translocating glycosyl transferases (4.D; [30]; [31]). Each of the eight different strains has between 7-12 such enzyme-transporters, with K12 having eight and UMN having only seven. Interestingly, the pathogens possess eight sets of orthologs not found in K12, suggesting that unique glycosyl transferases may serve different functions in the different pathogens. Some of them could function in the production of capsules, known to be virulence factors in many pathogenic bacteria [32].

Subclass 5.A includes paired electron-carriers that transfer two electron from one side of the membrane to the other, thereby influencing cellular energetics [33]. The organisms examined have between 42-45 such proteins, with K12 having the maximal number. Intriguingly, only 5 such proteins are non-orthologous to those found in K12 showing that the vast majority of these proteins are shared among the strains. Many of these oxidoreductases are involved in periplasmic protein sulfhydryl group oxidation and reduction using electrons from the cytoplasm or donating electrons to cytoplasmic receptors. Although these carriers have been implicated in defense against oxidative stress, it seems unlikely that they play a direct role in virulence [33].

These *E. coli* strains have far fewer transmembrane one electron carriers (5.B) than two electron carriers per organism; the numbers varying between 2-4. Since all of the proteins in this subclass are orthologous to the three K12 proteins, at least two of the paralogs found in UMN may have resulted from a recent gene duplication event.

Auxiliary proteins (8.A) vary in number between 33-45 with K12 having the lowest number. Twenty of the orthologs found in pathogens are lacking in K12. Surprisingly, all of the eight strains examined possess homologues of the related $b0^{+}$ amino acid transport (rBAT) accessory protein (TC#8.A.9), characterized only in eukaryotes. However, rBAT shows sequence similarity to various carbohydrases. These strains also have the 37 amino acid peptide toxin, BgK, the only protein belonging to subclass 8.B found in any *E. coli* strain.

Tremendous variation is observed between the numbers of poorly characterized transporters found in subclass 9.A. K12 has the fewest, with only 42 such proteins while 559 has over twice this number with 87. Most of the pathogens have between 60-70 such proteins, and 94 of these proteins are lacking in K12. This shows that the pathogens possess large variable numbers of poorly characterized transporters that are likely to serve highly specific functions within the cell. It is probable that they remain poorly characterized because they are not found in *E. coli* K12 and other commonly studied model bacteria. Elucidation of their functions and mechanisms of action is likely to prove particularly interesting and important for specific aspects of *E. coli* pathogenesis.

Subclass 9.B includes proteins that may function in transport, but the evidence for this suggestion is incomplete. Thus, not all of these systems are likely to prove to be transporters. In contrast to the situation observed for subclass 9.A, the strains examined have relatively little variation in the numbers of subclass 9B members. The smallest number is found in O15 (81), and the largest number is found in CFT (91) with K12 having 90. Nevertheless, the pathogens collectively have 26 such proteins lacking in K12.

Transport proteins likely to be important in pathogenicity include (1) toxins, (2) trans-envelope protein secretion systems, (3) outer membrane protein secretion systems, and (4) outer membrane iron-siderophore receptors that function with cytoplasmic membrane ABC-type iron uptake porters. The transport proteins that fall into these four classes are summarized in Tables 3 through 6.

Table 3 summarizes the toxins identified in the eight *E. coli* strains included in this study. In addition to the pore-forming subunits that function in conjunction with type III secretion systems (IIITCP), seven families of toxins are represented. These will be discussed sequentially according to their TC numbers.

HlyE is a well-characterized toxin found in *E. coli, Salmonella* and *Shigella* species. Its high resolution X-ray structure is available, and models of its membrane-inserted form have been proposed [34]. HlyE forms a homooligomeric pore complex where each subunit has 303 aas. The transmembrane pore has a diameter of 2.5-3.0 nm and oligomerizes primarily into octameric or dodecameric nonselective pores upon

binding to cholesterol in the host membrane [35]. Five of the eight *E. coli* strains examined possess this toxin. They include *E. coli* K12 as well as all three uropathogenic strains examined as well as the only enterohaemorrhagic strain, O15. This toxin was found to be lacking in the enteroaggregative (559), adherent invasive (APE), and enterotoxigenic (E24) strains.

Another haemolysin is the *E. coli* Haemolysin A (HlyA), part of the RTX superfamily of pore-forming proteins [36]. The general mechanism proposed for HlyA is as follows: the toxin (1) recognizes and binds to β 2-integrin receptors on host cell membranes, (2) inserts into the host membrane forming nonselective pores, and (3) transport a toxin that inhibits host protein kinase B, leading to apoptosis of the host cell [37]. This toxin was found in only two of the pathogenic strains, both of which have HlyE as well. They are ABU, which has two such toxins, and CFT, which has one. Both of these strains are intracellular pathogens.

Type 3 Secretion System (T3SS) pore-forming proteins (IIITCP) are typically involved in attaching/effacing mechanisms in enteropathogenic pathovars [16]. The EspB/D complex and other homologous pore-forming protein pairs are secreted into the host cell and oligomerize to form translocation pores at the site of contact. EspB with EspD, for example, provide translocation functions for *E. coli* effector proteins and inhibit host phagocytosis by altering host cell cytoskeletal functions [38]. Bacterial IIITCP family proteins were found specifically in two pathovars but were lacking in all others. These T3SS possessing strains are UMN (uropathogenic) and O15

(enterohaemorrhagic) [39]. The former strain encodes the two components of a IIITCP pore-forming complex that most closely resemble SipBD (1.C.36.3.2). As will be shown below, this organism has multiple components of a T3SS, consistent with its possession of the pore complex. O15 apparently has one pair of IIITCP pore-forming components, which most resemble EspBD (1.C.36.1.1), and it also has an EspA-like protein that is distantly related to SseB of *Salmonella typhimurium* (1.C.36.5.1). Thus, this strain possesses a single intact pore-forming complex consistent with its possession of a single T3SS.

The prophage-encoded shiga toxin (Stx) is a heteromeric protein with a glycolipid-recognizing, pore-forming, pentameric B subunit and a single A chain with N-glycosidase activity that results in cleavage of a single adenine in the 28S rRNA, resulting in blockage of protein synthesis [40]. Two StxB pore-forming toxins appear to be encoded in the O15 genome, although no such toxin was identified in the other strains. StxB components of these two toxins are 89 amino acids in length and differ from each other in only three positions, thus being nearly identical to each other.

Clostridial cytotoxins (CCT) are tripartite proteins with N-terminal catalytic domains , C-terminal receptor regions and central channel-forming domains [41]. The mechanism of action may be as follows: (1) Upon endocytosis into the host cell, the channel-forming domain is activated by acidic pH which induces conformational changes promoting insertion into the endosomal membrane, concomitant with channel formation. (2) The catalytic domain of CCT is translocated across the membrane. (3)

Autoproteolytic processing cleaves the glycosyltransferase moiety from the remainder of the protein, releasing it into the host cytoplasm. (4) The glycosyltransferase-dependent inactivation of Rho family GTPases in the host cell leads to reorganization of the host actin cytoskeleton [42]. Two homologs of the clostridial cytotoxin (TC #1.C.57.3.2 and 3.3) were identified, one in ABU (UPEC) and one in O15 (EHEC). These proteins are 85% identical to each other and are probably not orthologous.

Serratia-type pore-forming-type toxins (S-PFT) (1.C.75), originally identified in *Serratia marcescens,* are exported through a two partner secretion system (TPS; TC # 1.B.20) [43]. Three organisms possess this type of pore-forming haemolysin, ABU (UPEC), CFT (UPEC) and O15 (EHEC). The former two have huge toxins of 3216 amino acids that are nearly identical and are clearly orthologous. By contrast, the O15 homolog is 1270 amino acids in length and corresponds in sequence to the N-terminal portion of the larger toxins. This protein is also sequence divergent relative to the other two homologs.

Not surprisingly, cytotoxic major fimbrial subunits (TC#1.C.80) were identified. Some of these have been reported to be capable of pore formation [44]. However, the physiological significance of this observation is not known.

Toxins of the Cytolethal Distending Toxin (CDT) family are tripartite protein complexes, CdtABC, that induce DNA cleavage, cell cycle arrest, and apoptosis in host epithelial and immune cells [45]. The catalytic subunit of the complex is CdtB, which

possesses phosphodiesterase activity that cleaves the host DNA, inducing double-strand breaks and causing G2/M cell cycle arrest in the host cell [46]. CdtA and C are dimeric subunits that deliver CdtB to the host cell cytoplasm [47]. The CdtA/B/C trimeric toxin system (1.C.98.1.1) was found in only one of the eight *E. coli* strains, APE (AIEC) where all three subunits were identified.

Table 4 summarizes the various protein secretion systems found in the *E. coli* strains examined. These include (1) type I secretion systems (T1SS; ABC) (TC#3.A.1), (2) type II secretion systems (T2SS; SEC-SRP complexes) (TC#3.A.5) for secretion across the inner membrane, acting together with the Main Terminal Branch (MTB; TC# 3.A.15) for secretion across the outer membrane, (3) type III secretion systems (T3SS; Fla; Path; TC #3.A.6), (4) type IV secretion systems (T4SS; Conj; Vir; TC# 3.A.7), and (5) the relatively recently characterized type VI secretion systems (T6SS; TC# 3.A.23) [48]. In previous studies, T3SS, T4SS and T6SS have been shown to contribute to the pathogenicity of specific *E. coli* strains by mechanisms involving injection of effectors into target cells [16].

Type I secretion systems (T1SSs) identified are members of the ABC Protein Exporter-1 (TC#3.A.1.109) and Protein Exporter-2 (TC#3.A.1.110) families within the ABC functional superfamily [49], which often catalyze export of pathogenicity proteins. ABC exporters consist of two transmembrane domains and two cytosolic ATP-hydrolyzing domains. The current paradigm for T1SS efflux involves ATP hydrolysis-coupled translocation of the substrate across the two bacterial membranes in a single

energy-coupled step, utilizing an outer membrane factor (OMF; TC#1.B.17) and a membrane fusion protein (MFS; TC#8.A.1) to couple cytoplasmic membrane transport with that across the periplasm and outer membrane ([50], [51]). Protein exporter-1 family proteins were found in four of the pathovars examined in our study (ABU, CFT, UMN and O15). The α-hemolysin exporter, HlyB (TC #3.A.1.109.1), is found in ABU (UPEC) and CFT (UPEC) only. One complete system of the LapA adhesion protein exporter (TC #3.A.1.109.3) and the biofilm exporter (TC #3.A.1.109.4) are present in each of the four pathovars mentioned above. Only the ABU and CFT have a toxin exporter of the MceGH-type (3.A.1.110.1), but a function in pathogenesis for Microcin E492, the substrate of MceGH, has not been established. ABC peptide secretion systems are present in all strains (Table 4).

Type III secretion systems are found in numerous Gram-negative bacterial strains and are used for secretion of flagellar and pathogenicity effector proteins. Effector proteins can be exported across both bacterial membranes as well as the host cell membrane in a single step. The type III protein secretion system includes a needle complex known as the injectisome, which is embedded in the bacterial inner membrane, crosses the outer membrane, and attaches to the host cell membrane via the two constituents of the T3SS pore-forming complex discussed above (e.g., EspB and EspD) [52], creating a channel in the host cell membrane [53]. Effectors of Ysc T3SS homologs influence host inflammatory responses and cell death (TC #3.A.6.1). ABU (UPEC), APE (AIEC), CFT (UPEC) and the non-pathogenic K12 contain homologs of the energizer protein, YscN, but not other constituents of these systems ([54] ; [55]). The 559 (EAEC)

and E24 (ETEC) strains appear to contain more complete, but still only partial systems with four and six constituent proteins, respectively, of the twenty-two known protein subunits in the Ysc T3SS. Only the UMN (UPEC) and O15 (EHEC) strains appear to contain all or most of the essential transmembrane subunits of the Ysc system, including LcrD (7 TMSs), YscU (5 TMSs), YscJ (4 TMSs) and YscT (7 TMSs) ([56] ; [39]; [55]).

As expected, the complete general secretory pathway (TC#2.A.5.1.1) and the complete flagellar protein export complex (TC #3.A.6.2.1) were found in all strains. The UMN (UPEC) pathovar appears to possess six additional key components of the flagellar system (FliI, FliN, FlhB, FliP, FliQ, and FliR) (Table 4).

T3SSs and T4SSs allow direct contact with the host cell for transport into the recipient cell cytoplasm. There are three general types of T4SSs: (1) conjugation systems, (2) host cell effector protein translocation systems, and (3) DNA release systems for export into the extracellular medium. The VirB system of *Agrobacterium tumefaciens* (TC#3.A.7.1.1) embodies the paradigm of T4SS transport, which manifests its structure in various other systems in the subfamily. The VirB11 ATPase, VirB4, VirB6, VirB7, VirB8 and VirB9 complete the secretion channel and can be cross-linked to DNA upon secretion. VirB3, VirB4 and VirB10 assist with the formation of the protein channel [57]. However, T4SSs vary tremendously with respect to sequence divergence and numbers of constituents required for function. An examination of potential T4SS components in the

eight *E. coli* strains revealed the presence of 1-5 potential subunits, suggesting that some of these organisms may utilize this mechanism to promote pathogenesis.

In our study, proteins homologous to those in two β-proteobacterial T4SS were identified as recorded in Table 4. The first type is a DNA secretion system (TC #3.A.7.11.1), and only APE has as many as four constituents. Four pertussis toxin exporter component homologs (TC #3.A.7.3.1) were identified only in UMN (UPEC). An essential subunit, a PtlH homolog, which is required for transporting pertussis toxin across the outer membrane, was identified in UMN. Other subunits, such as PtlE (aka VirB8), PtlF (aka VirB9 OM component), PtlH (aka Vir11 ATPase), and IVSP PtlC were also identified in the UMN strain, indicating the presence of both IM and OM complexes [58]. We therefore suggest that this one strain has the equivalent of the pertussis toxin export system.

The more recently characterized Type VI secretion systems are found in multiple Gram-negative bacteria, including *Vibrio cholerae* and *Legionella pneumophila*. Key constituents of T6SSs include an IM ATPase protein, ClpV, a phage tail-like protein that spans the extracellular matrix, and a "tail-spike" protein, VgrG, that penetrates the host membrane. After contacting the host membrane, the tail-spike dissociates from the complex to allow contact-dependent translocation of proteins into the host cytoplasm [59]. In the *E. coli* strains studied here, two sets of complete T6SS complex homologs were identified (VasA-L of 3.A.23.1.1 and EvpA-P of 3.A.23.2.1). For the first of these two systems, at least seven out of the fourteen subunits were found in each of the

pathovars studied. 559 (EAEC), APE (AIEC), E24 (ETEC), UMN (UPEC) and O15 (EHEC) all contain the necessary components of a T6SS, including the IcmF energizing protein, a ClpB protein, and Vgr-3 (Table S1). ABU and CFT, both UPEC strains, possess only partial systems with eight and seven out of the 14 protein components, respectively. Only 559 contains a full EVP T6SS. Pathovars ABU, APE, CFT and E24 have only partial systems with five or six proteins out of the 16 components. UMN and O15 possess only two of 16 proteins each. A homologue of only one component (EspH/ClpB, an ATP-binding protein) was found in K12. Thus, several of the pathogens appear to have complete type 6 secretion systems.

Table 5 summarizes the outer membrane protein secretion systems present in the eight *E. coli* strains studied. These include the autotransporter families, Autotransporter-1 (TC#1.B.12), Autotransporter-2 (TC#1.B.40), outer membrane factors (OMF; TC#1.B.17), fimbrial usher proteins (FUP; TC#1.B.11), two partner secretion systems (TPS; TC#1.B.20), secretins (TC#1.B.22), outer membrane protein insertion porins (OmpIP; TC#1.B.33), curli fiber subunits (CsgA; TC#1.B.48), and putative autotransporter-3 (Invasins; TC#1.B.54). Autotransporters are virulence factors that insert into the outer bacterial membrane to form transmembrane β-barrels which export their extracellular protein domains. Autotransporter-1 proteins each consists of an N-terminal cleavable secretory signal, an exported passenger domain of variable lengths, and a C-terminal 250-300 amino acyl residue domain that inserts into the outer membrane, giving rise to a 12 TMS β-barrel structure [60]. Upon translocation of the N-terminal residues across the outer membrane, proteolytic cleavage usually occurs in which conserved

aspartyl and asparaginyl residues of the C-terminal β-barrel domain fold into a catalytic dyad that catalyzes passenger domain cleavage, thereby releasing the N-terminal effector protein [1].

Autotransporter adhesins (e.g., AidA; TC#1.B.12.1.1) appear to be present in all eight *E. coli* strains examined (Table 5), but virulence factor IcsA (TC#1.B.12.1.2) is lacking in two of the pathovars as well as K12. Two such non-orthologous virulence factors occur in 559, ABU, CFT and E24, but just one is present in O15. Fibronectin binding proteins (TC#1.B.12.1.3) assist in adhesion to host cells and were identified in all pathovars except ABU. An additional fibronectin binding protein autotransporter was identified in UMN, but none was found in K12. A pertactin autotransporter homolog (TC#1.B.12.2.1) was found in 559, E24, UMN and K12. A close homolog, a tracheal colonization factor autotransporter (TC#1.B.12.2.2) was absent in K12 and O15 but present in all other strains. Another pertactin autotransporter, BrkA (TC#1.B.12.2.3), was identified only in APE. A putative autotransporter that is homologous to one in a dual membrane firmicute autotransporter (TC#1.B.12.2.4) was identified in several pathogenic strains. Three non-orthologous homologs of such a protein were found in ABU and CFT, two were found in 559, and one was found in APE and UMN. *Salmonella* BigA autotransporter (TC#1.B.12.5.5; [61]) homologs were identified in three strains: ABU, APE and UMN. Autoaggregation factor (TC#1.B.12.8.2) was found in all strains but was present in greater numbers in ABU, CFT and E24. Autotransporters for the adhesin, TibA, were found in all strains, but in larger numbers in 559, ABU, APE, E24, (2 copies each) and in UMN and O15 (three copies each) (Table 5). Homologues of EspP

(TC#1.B.12.4.3) a recognized O15 virulence factor [62], and temperature-sensitive hemagglutinin (Tsh; TC# 1.B.12.4.2) were found in strains 559, ABU, APE, CFT, UMN and O15. These virulence factors were found in greater numbers in 559, ABU and CFT, where three non-orthologous homologs are present. Overall, the most autotransporter-1 proteins were found in ABU (UPEC) with sixteen homologs; fourteen were found in 559 (EAEC), twelve in CFT (UPEC) and UMN (UPEC), ten in APE (AIEC) and E24 (ETEC), and eight in O15 (EHEC). K12 contains the fewest with only five.

Autotransporter-2 family proteins have trimeric structures organized into three domains: an N-terminal head that adheres to the host cell membrane, a stalk, and a C-terminal anchor, rich in glycine, which forms a β-strand domain that oligomerizes to form a pore for auto-transport [63]. Haemagglutinins (TC#1.B.40.1.2) were identified in only two of the pathovars: in UMN (UPEC) and O15 (EHEC). A dissimilar adhesin, YadB (TC#1.B.40.1.3) [64] was found only in O15. A cryptic trimeric *Haemophilus* adhesin homolog (TC#1.B.40.1.5) was also found in UPEC strain UMN as well as in APE (AIEC). The trimeric UpaG (TC#1.B.40.2.3) was found in all serovars except UMN and K12. Interestingly, AT-2 family proteins are far less prevalent than AT-1 family proteins in all strains and were lacking in K12.

Invasins or intimins are also called Autotransporters-3, but a function in autotransport is not well established [65]. The N-terminal domain serves as an anchor and inserts a pore-like β-barrel in the outer membrane. The C-terminus contains folds that resemble C-type lectin-like receptor binding domains [66]. These domains bind to Tir

(translocated intimin receptor) and β-1 integrins on host cells, leading to pathogenesis in enterohaemorrhagic and enteropathogenic strains [67]. In our study, intimin/invasin proteins were identified across all serotypes but were found in abundance in six of the seven pathogens. The UPEC strains have more intimins/invasins than the enteropathogenic strains.

Outer membrane factors (OMFs; 1.B.17) and membrane fusion proteins (MFPs; TC# 8.A.1) function with Major Facilitator Superfamily (MFS) type porters, Resistance-Nodulation Cell-Division [68] family proteins (2.A.6), ATP-Binding Cassette (ABC) porters and aromatic acid exporters (AAE, TC#2.A.85) to form complexes that export various drugs, oligosaccharides, proteins, aromatic acids and siderophores across both the inner and outer membranes in a single energy-coupled step ([69], [70], [71], [72], [73]). OMF proteins form homotrimeric, 12-stranded β-barrel complexes that form holes in the outer membrane for translocation. The OMF component, TolC (TC #1.B.17.1.1), was identified in all eight strains. TolC-dependent exporters translocate hemolysins, drugs and various siderophores [74]. An additional TolC-like protein, MdtP, was found in each pathogen but not in K12. The MdtP OMF acts with the MdtO drug exporter (2.A.85.6.1) and auxiliary MFP protein MdtN (8.A.1.1.3) in multidrug resistance [75].

The Fimbrial Usher Proteins of the FUP family function in conjunction with periplasmic chaperone proteins in fimbrial assembly [76]. The usher proteins contain central domains that span the outer membrane 24 times to form β-barrel pores. The C-terminal domains act as “plugs” that lie within the lumen of the β-barrels [77]. These plug

domains, along with helical folds, regulate channel opening and export of proteins that contribute to fimbrial assembly [78]. Type π FUPs (1.B.11.2; [79]) were identified in all strains of *E. coli* but are observed in greater numbers in the pathovars. With the exception of UMN, each pathovar possesses at least three types π FUPs. Type α FUPs (1.B.11.4) were also identified. 559 (EAEC) and E24 (ETEC) strains possess two such proteins while other serotypes possess only one. Type γ FUPs were also identified in the various strains, and quantitative differences were observed among the serotypes.

Homologs of a single two-partner secretion (TPS) system (ShlB; TC#1.B.20.1.1) were identified in this study. Transport of ShlA hemolysin across the cytoplasmic membrane in *S. marcescens* involves the General Secretory Pathway (GSP, TC #3.A.5). The oligomeric β-barrel channel of ShlB serves in secretion of the effector protein across the outer membrane [80]. These TPS systems were found exclusively in O15 (EHEC), where two paralogues were present, and in ABU (UPEC), where a single protein was identified.

A Secretin (TC#1.B.22) consists of a 280aa N-terminal periplasmic domain and a C-terminal domain responsible for forming large channels via homomultimeric ring structures [81]. The C-terminus interacts with both the peptidoglycan on the periplasmic side and the lipopolysaccharide on the cellular surface. Secretins participate in various transport processes such as fimbrial protein export, T2SS, T3SS, and phage export ([82]; [83]). In our study, YscC-like secretins were found exclusively in UMN and O15. A homolog of the bacteriophage Gene IV protein (1.B.22.5.1) was found only in E24

(ETEC). HofQ (1.B.22.4.2), possibly involved in DNA uptake in some *E. coli* strains, although several of these strains including K12 have been shown to lack natural competence [84], was found in all strains. Similarly, the PulD secretin homolog (1.B.22.1.1) was present in seven of the eight strains although not all of these strains (e.g., K12) have a complete outer membrane MTB protein translocase (3.A.15) [52]. As expected, outer membrane protein (OMP) insertion porins (OmpIP or BAM, TC #1.B.33), required for OMP insertion, were ubiquitously present, and all six required subunits were identified in all eight strains. Multiple curli fiber subunit export proteins, CsgG (TC #1.B.48), were identified in all strains.

Table 6 lists outer membrane receptors of TC class 1.B.14, most of which are involved in the uptake of iron-siderophore complexes. This table also includes a cobalamin receptor and a copper receptor and indicates the occurance of TonB and TolA proteins involved in OMR energization. For some of these proteins, a single ortholog was found in all eight organisms. For example, the following receptors have this characteristic: TC# 1.B.14.1.1 (ferric copragen), 1.B.14.1.2 (ferrichrome), 1.B.14.1.9 (catecholate), and 1.B.14.3.1 (cobalamin). Some organisms possess more than one gene encoding a specific type of iron complex uptake receptor. These include: 1.B.14.1.4 (catecholate) where four of the organisms have a single gene encoding this system, but three organisms have two, and one (CFT) has three (Table 6). A similar situation was observed for 1.B.14.7.2, an iron pesticin uptake receptor. In this case, four of the organisms including K12 have a single gene encoding such a system, while four of the pathovars have two. Six additional types of outer membrane receptors specific for iron

complexes were found in some of these *E. coli* strains but not others. These include a ferroxamine receptor (TC#1.B.14.1.10), the ferric-citrate receptor (TC#1.B.14.1.12), CfrA ferric receptor homologs (TC#1.B.14.1.13), a dissimilar ferrichrome receptor (TC#1.B.14.1.15), a heme receptor (TC#1.B.14.2.2), and an aerobactin receptor (TC#1.B.14.9.3). Only two of the UPEC pathovars (ABU and CFT) possess a ferric-enterobactin receptor (TC#1.B.14.1.3) and a ferrichrome receptor (TC#1.B.14.1.15), although in the case of most other iron complex receptors, a larger selection of pathovars possesses the system. For example, four or five of them possess most of the remaining iron complex receptors, but each receptor type was found in a different set of these organisms. Similarly, only four of the pathovars possess the copper receptor (TC#1.B.14.4.1). All in all, each of these organisms possesses nine to fifteen iron uptake receptors except for K12 and enterotoxigenic E24, which have only six. Interestingly, the uropathogenic strains all possess more (ten to fifteen) iron complex receptors than the other serovars.

Table 6 also includes the energizing modules, TonB and TolA of family 2.C.1 [85]. It can be seen that while *E. coli* K12 has just one of each of these two proteins, four of the pathogens have two TonB's, five of them have two TolA's, and one (559) has three TolA's. Each of these energizers probably interacts with a different set of outer membrane receptors.

Table 7 summarizes inner membrane iron uptake systems identified in the eight *E. coli* strains. These systems function to transport iron from the periplasm into the

cytoplasm of the bacterium, complementing the outer membrane iron-scavenging OMR complexes, which import iron from the surrounding extracellular environment into the periplasm. Various mechanisms are employed for inner membrane transport of iron/iron-chelate complexes. These include the CorA heavy metal ion uptake channels (1.A.35), the zinc-iron permeases (ZIP (2.A.5)), metal ion transporters (Nramp (2.A.55)), GTPase-coupled ferrous iron uptake systems (FeoB (9.A.8)), and ATP-binding Cassette (ABC (3.A.1)) transporters, all of which were examined.

Zinc-Iron Porters (ZIP) have an average of eight TMSs with histidine-rich repeats near the N- and C-termini [86]; the N-termini have been implicated in metal selectivity [87]. Mutations in ZIP transporters in *Salmonella* affect the specificity and kinetics of metal influx [88]. A cation symport mechanism may be operative [89]. A single such transport protein was found in each of the eight strains examined (Table 7).

Metal ion transporter (Nramp) homologues are found ubiquitously. These proteins generally function by a symport mechanism in a 1:1 ratio of ferrous ion and H^+ although stoichiometrics have been reported to be variable [90]. Homologues of Nramp proteins usually possess 11-13 TMSs. A single transport protein was found in each of the eight strains of *E. coli* examined (Table 7).

Another iron influx system belongs to the iron uptake family, FeoB (TC#9.A.8). These transport proteins possess on average 10-12 TMSs. Their N-termini contain hydrophilic catalytic domains with sequences homologous to GTP-binding domains (e.g.,

Era protein of *E. coli*) [91]. Their C-termini catalyze transport, possibly coupled to K^+-activated GTPase activity. GTP hydrolysis has been reported to be required for transport, and if there is direct energy coupling, the mechanism may resemble that of ABC transporters ([92]; [93]; [94]; [95]). Two FeoB homologues were identified in each of the eight strains of *E. coli*.

ATP-binding cassette (ABC) porters couple ATP hydrolysis to transport. Four types of such inner membrane iron-uptake systems were identified: ferric uptake systems (TC#3.A.1.10), iron chelate uptake transporters (TC#3.A.1.14), manganese/zinc/iron chelate uptake transporters (TC#3.A.1.15), and homologs of the *Brachyspira* iron transporters (TC#3.A.1.20). Interestingly, a single homologue of a ferric uptake system (TC#3.A.1.10) was identified in each of the various strains of *E. coli*, but none of the eight strains appears to possess a complete system. A complete four-subunit ferric-citrate uptake system (TC#3.A.1.14.1) was identified in five of the eight strains, including 559 (EAEC), ABU (UPEC), E24 (ETEC), UMN (UPEC) as well as K12. Complete ferric-enterobactin (TC#3.A.1.14.2) uptake and ferric-hydroxamate (TC#3.A.1.14.3) uptake systems were identified in all eight strains of *E. coli* examined. However, a second complete ferric-enterobactin system (TC#3.A.1.14.2) was identified in ABU (UPEC) and CFT (UPEC).

An interesting case was observed for iron-vibrioferrin uptake systems (TC#3.A.1.14.8), where ABU (UPEC), APE (AIEC), CFT (UPEC), UMN (UPEC) and O15 (EHEC) all possess two partial systems. Homologs of the *Shigella* heme uptake

system (ShuTUV (3.A.1.14.18)) were present in ABU, APE, CFT, UMN and O15 strains as complete systems. Complete systems corresponding to the *Salmonella* Mn^{2+} and Fe^{2+} uptake systems (3.A.1.15.7) were identified in ABU, APE, CFT and UMN. Three subunits (BitA (Q9Z4R9), BitF (O54372) and BitD (O54370)) of the six-subunit iron transporter of *Brachyspira* (3.A.1.20.1) were found exclusively in O15. BitD is the ATP-binding domain, BitF is one of the two integral membrane domains, and BitA is one of the three periplasmic binding proteins of the *Brachyspira* system. The presence of all three types of subunits indicates that a complete Bit system may be present.

Table 8 presents an overview of the total number of iron-siderophore transport complexes in each strain, including both outer membrane receptors and inner membrane transporters. The juxtaposition of the two data sets demonstrated that the UPEC strains, ABU and CFT, contain the greatest number of both OMRs and inner membrane iron-siderophore transporters. E24 (ETEC) and K12 both contain the smallest numbers of both iron-siderophore OMRs and inner membrane iron transport proteins with six OMRs and 16 porter constituents, respectively. 559 (EAEC), APE (AIEC), UMN (UPEC) and O15 (EHEC) each have intermediate numbers with 9 or 10 OMRs and 16-21 inner membrane transport proteins.

Figure 3 reveals the correlation between numbers of OMRs and numbers of iron ABC uptake systems. A clear correlation between the number of iron-siderophore OMRs and inner membrane transporters is evident for the eight strains of *E. coli* (RSQ value =

0.90). The slope of the trendline (SLOPE = 1.85) confirms the positive correlation between the two sets of values.

We identified numerous nutrient uptake porters that are uniquely found only in the three intracellular pathogens or only in the extracellular pathogens (Table 9). Transporters found exclusively in the intracellular pathovars include certain α-type channels (TC#1.A), β-barrel porins (1.B), holins (1.E), MFS porters (2.A.1), multidrug resistance (MDR) RND-type efflux pumps (2.A.6), ABC porters (3.A.1), PTS sugar uptake transporters (4.A), and various TC class 9.B proteins. Thus, the APE, CFT and ABU strains, all primarily intracellular pathogens, possess many transport systems that take up metabolic products produced via host intracellular metabolic processes. Examples include a phosphoglycerate:Pi antiporter (2.A.1.4.2) that takes up glycolytic intermediates, 2- and 3-phosphoglycerates as well as phosphoenolpyruvate, a pmf-dependent acetate uptake system (TC#1.A.14.2.2), an ATP-dependent ribose porter (3.A.1.2.1) and a phosphoenolpyruvate (PEP)-dependent D-tagatose/D-psicose/D-sorbose PTS uptake system (4.A.2.1.9). While ribose is a substrate of RNA biosynthesis and a product of RNA degradation, the latter 2-keto hexoses are intermediates of several intracellular metabolic pathways and of nutritional value ([96]; [97]). These intracellular pathogens also uniquely possess a citrate:H^+ symporter (2.A.1.6.1), a fumarate-malate transporter (2.A.13.1.2), a citrate:succinate antiporter (2.A.47.3.2), a tripartite tricarboxylate (citrate, isocitrate and cis-aconitate) transporter (2.A.80.1.1), and a putative C4-dicarboxylate transporter (9.B.50.1.2), all of which are specific for intermediates in the Krebs (TCA) cycle, and none of which are found in the extracellular pathovars. In

addition to Kreb cycle metabolism, C4-dicarboxylates can be generated via multiple pathways including amino acid deamination and lipolysis.

Two of the intracellular pathogens, APE and CFT, possess an arginine/agmatine antiporter (2.A.3.2.7), which allows these cells to survive under acidic conditions, for example, as exist inside lysosomes [98]. These intracellular pathogens also possess a probable antimonite/arsenite resistance exporter (2.A.47.4.2). Arsenite can be passively taken up by host cells via aquaglyceroporins (TC # 1.A.8) [99]. Additionally, an ATP-dependent nucleoside group translocator (4.B.1.1.1) that synthesizes NMN, an intermediate in NAD synthesis, and a thiamin porter (4.B.1.1.4), involved in the synthesis of thiamin pyrophosphate (TPP), were found in these strains but not in the extracellular pathogens (Table 9). While NAD is a common cofactor of numerous dehydrogenases, TPP participates in decarboxylation reactions such as occur in the pentose phosphate pathway during aerobic sugar metabolism. Their precursors, NMN and thiamin, are accumulated intracellularly over 1000-fold relative to the extracellular concentrations.

Surprisingly, the intracellular pathogens possess a glucuronate/galacturonate porter (2.A.1.14.2), which is absent in the extracellular pathogens (Table 9). These metabolic products have been shown to be generated intracellularly when exogenous sugars such as lactose are metabolized by *E. coli* [100]. They may play a protective role under osmotic stress conditions. These compounds have been shown to be present intracellularly in the intestines, spleen, liver, and numerous other organs of animals, and these *E. coli* strains utilize cytoplasmic UPD-glucuronyltransferases for

mucopolysaccharide biosynthesis [101]. Additionally, a sialic acid uptake porter (2.A.56.1.6) is found exclusively in the intracellular pathogens. Although sialic acids are found in abundance in cell surface glycolipids and extracellular glycoproteins, *N*-acetylneuraminic acid, the main mammalian sialic acid, can be found intracellularly where it participates in glycolipid and glycoprotein biosynthesis [102]. Sialic acid-containing polysaccharide capsules are virulence factors in *E. coli* urinary tract infections and contribute to biofilm formation [103].

Intracellular *E. coli* strains lack various transporters present in extracellular pathogens that take up metabolites normally found extracellularly. For example, the latter, but not the former strains possess a rhamnose oligosaccharide porter (3.A.1.5.12), a melibiose permease (2.A.2.1.1), and a 2-O-α-mannosyl-D-glycerate porter (4.A.2.1.3). These first two oligosaccharides are constituents of glycolipid and glycoproteins found on the external surfaces of intestinal villus cells, while the last compound is a known intracellular osmolyte in some bacteria [104]. The extracellular strains, but not the intracellular strains, also have a xylose:H^+ symporter (2.A.1.1.3), a mannose PTS porter (4.A.2.1.6), a glucitol PTS transporter (4.A.4.1.2), and a galactitol PTS permease (4.A.5.1.3). The former two sugars are found in abundance in cell surface polysaccharides ([105],[106]) while the hexitols are common osmolytes in animals [107]. The α-glucoside-specific IICB permease of the PTS (4.A.1.1.10), which transports many glucosides including methyl-α-glucoside, maltitol, isomaltose, trehalose, turanose, maltulose, leucrose and palatinose, is similarly found in the extracellular pathogens but lacking in intracellular strains as makes teleological sense (Table 9).

Other transporters lacking in intracellular but present in extracellular pathogens include an uptake system for the folate catabolite, p-aminobenzoyl-glutamate (2.A.68.1.1). APE, CFT and ABU also lack the cyanate porter (2.A.1.17.1), which transports toxic cyanate and thiocyanate that can be used by *E. coli* strains as nitrogen sources for growth if the *cyn* operon that includes cyanase and the cyanate permease is present ([108] ; [109]). Thiocyanate and cyanate are found in extracellular environments and are taken up via the CynT permease for bicarbonate-dependent destruction by cyanase, CynS [108]. *E. coli* cells lacking the *cyn* operon are more sensitive to the toxic effects of cyanate and thiocyanate than strains that possess it. Both mammals and bacteria have transporters for these compounds [110].

A putrescine uptake porter (2.A.3.1.13) was found to be present only in extracellular strains. Putrescine, like other positively charged polyamines, stabilizes the negatively charged DNA during replication, but it is also found in abundance in extracellular environments in mammals such as seminal fluid and urine, especially during episodes of protein breakdown ([111] ; [112]). Several MDR and iron-siderophore transporters are found exclusively in intracellular pathogens, while others are present in extracellular pathogens exclusively. These facts may reflect the natural cellular substrates of MDR pumps currently not recognized. The need for specific types of iron uptake systems may reflect the need for high versus low affinity iron-specific or iron complex-specific transporters in the different host organismal sites, intra- versus extracellular, as discussed above.

Several proteins were identified that gave poor scores in the G-Blast search, but which nevertheless proved to be homologous to the TC protein recognized. These are presented here in order of their TC numbers. (1) A glutathione-S-transferase homolog was identified which showed sequence similarity to TC #1.A.12.3.1. In animals, these proteins can insert into membranes to form chloride channels of the CLIC family. There is no evidence that these proteins can form channels in bacteria. (2) A protein was identified that exhibits poor similarity to the *Bacillus* gap junction-like channel-forming complex protein, SpoIIQ. In fact, the four homologs gave poor scores to the *Bacillus* protein. However, no homolog of the *Bacillus* partner protein, SpoIIIAH, was identified, leaving it uncertain as to whether the SpoIIQ homologs in *E. coli* function in a similar capacity. (3) A putative invasin with similarily to proteins in the OmpA family (TC#1.B.6) was identified and given the TC number 1.B.6.2.13. (4) A homolog of pore-forming phage coat protein A of the same topology and length as 1.B.53.1.1 was identified and assigned TC# 1.B.53.1.4. (5) Several proteins, including one designated SfaD (F1c fimbrial subunit of 179 aas), proved to be homologous to the cytotoxic pore-forming oligomeric fimbrial subunit, MrxA (TC#1.C.80.1.1). One such protein has been entered into TCDB under TC#1.C.80.1.2. (6) A holin of 108 aas was identified and entered into TCDB with TC#1.E.2.1.10. This protein was encoded by a gene within a prophage segment of the genome of UMN with two adjacent genes, a lysozyme inhibitor, and a lytic protein, probably a lysozyme. This observation strengthens the conclusion that this protein is a holin, probably of phage origin. (7) An ABC porter, YnjBCD, was assigned TC#3.A.1.19.4. Although it is annotated as a sulfate/thiosulfate transporter,

other members of this family transport vitamins. (8) Another protein designated YjiG has 4 TMSs with 153 aas and proved to be a distant member of the FeoB family (TC#9.A.8). Proteins of this family already in TCDB generally have a 9-12 TMS topology, but YjiG and all of its close homologs have about ~153 aas and 4 TMSs. More distant homologs proved to have 5 TMSs, and only the most distant homologs proved to have 9-12 TMSs. It seems likely that the larger proteins arose by duplication of a basic 4-5 TMS repeat unit. On the basis of annotations, a nucleoside-recognition domain may be present in some of these homologs. YjiG was entered into TCDB with TC# 9.B.156.1.1. (9) Additionally, two integral membrane serine proteases of the rhomboid family were identified. They have 625 aas with seven N-terminal TMSs, and one has been entered into TCDB (TC# 9.B.109.1.4). (10) The last poorly scoring protein proved to be a homolog of the DedA family. This protein contains 142 aas and 4 TMSs and was entered into TCDB with TC#9.B.27.3.1.

EHEC strain, O15, exhibited a number of characteristics that distinguished it from the other six pathogenic strains of *E. coli* examined. These will be briefly summarized here. O15 has two-partner secretion (TPS) systems corresponding to TC# 1.B.20.1.3 and 1.B.20.1.4. The former system plays a role in contact inhibition of growth [68] while the latter protein is a ShlB-type protein that contains a soluble polypeptide-transport-associated (POTRA) domain critical for substrate recognition and folding [113]. It probably secretes a haemolysin-like virulence factor [114]. Because this protein is distantly related to any proteins in TCDB, it has been entered into the database under

1.B.20.1.7. In contrast to the other strains studied, O15 lacks porins specific for chitosugars (TC# 1.B.25.1.5 and 1.B.25.1.13).

This strain uniquely has an AT-2 type adhesin, most closely related to TC# 1.B.40.1.3, as well as an intimin, most closely related to TC# 1.B.54.1.1. These proteins were found in no other *E. coli* strain examined. Most surprisingly, we identified 14 hits of 1.B.6.2.1 and 1.B.6.2.2, all related to OmpX of the OmpA family. These proteins are in the Ail/Lom family, which in *Yersinia* species are multifunctional virulence factors that protect against complement [115]. Equally surprising was the unique discovery of 12 putative holins corresponding to TC# 1.E.1.1.3. These holins may be related to phage-encoded shiga-toxins [116].

DISCUSSION

E. coli is a facultative anaerobe that utilizes iron primarily for electron transport and cellular respiration. A deficiency for iron leads to inhibition of the Kreb cycle that sustains electron transport. For example, succinate dehydrogenase and various cytochromes as well as NADH and ferric oxidase activities are diminished with an iron deficiency [117].

In humans, iron sequestration plays an important role in host defense against pathogens. Previous studies have evaluated roles of certain conserved eukaryotic genes that encode siderophore/iron-uptake complexes that allow the host to upregulate iron sequestration in response to inflammatory signals accompanying infection [118]. Interestingly, some of these uptake and iron-sequestration genes (e.g., Nramp1 and ferroportin) play pivotal roles in host defense in Crohn's disease, for their upregulation allows the host intestinal epithelial cells to sequester more iron through hepcidin transport [119].

In our study, the pathogenic *E. coli* strains examined proved to have greater numbers of iron-siderophore receptor-encoding genes than the nonpathogenic K12 strain. 559 (EAEC), APE (AIEC) and O15 (EHEC) strains all possess 9 unique iron-siderophore receptors while the UPEC strains contain 10-15 such proteins. These systems greatly outnumber those in K12, which has only 6 such receptors. Additional genes encoding siderophore biosynthesis and export as well as Fe^{3+}-siderophore complex ABC uptake

systems may contribute to pathogenicity by increasing gene dosage or by being expressed under different conditions.

Though certain siderophore receptors are ubiquitous among all strains (e.g., ferric coprogen, catecholate siderophores and ferric yersiniabactin uptake receptors), others were present in abundance in the pathovars but lacking in K12 (Tables 7 and 8). Such systems include a Fe^{3+} catecholate receptor (1.B.14.1.4) and the ferric yersiniabactin uptake receptor (1.B.14.7.2). The ferric yersiniabactin uptake receptor (1.B.14.7.2) is known to be required for biofilm production in UPEC strains [120], possibly relevant to host inflammatory signaling. Ferri-enterobactin receptors (1.B.14.1.3) and ferrichrome receptors (1.B.14.1.15) were identified exclusively in UPEC strains (ABU and CFT). These receptors may serve specific functions in UPEC pathogenicity. Further studies can be instigated to elucidate their involvement in pathogenicity.

All in all, the increased numbers of iron-complex receptors in these pathovars indicate targets for combating *E. coli* pathogenicity. Increases in siderophore receptor complexes may serve as compensatory mechanisms in response to host defense. Increased ferrichrome sequestration undoubtedly allows these pathovars to take up iron more readily, leading to increased survival in the host, thus sustaining invasion.

As shown in Table 8 and Figure 2, increases in numbers of iron-sequestering outer membrane receptors (OMR) correlate with complementary increases in inner membrane iron ABC transport systems. The latter transporters are essential in Gram-

negative bacteria for taking up the sequestered iron from the periplasmic space into the cytoplasm. We therefore examined iron transporters in the eight *E. coli* strains to evaluate the statistical significance of this correlation.

A strong positive correlation (RSQ = 0.90 with SLOPE = 1.85) was observed between the numbers of iron inner membrane ABC transport proteins and iron-siderophore outer membrane receptors as shown in Figure 2. In both cases, we found that the UPEC strains (ABU, CFT and UMN) have significantly larger numbers of both iron-siderophore OMRs and inner membrane transporters. We speculate that because UPEC strains must survive in both the gastrointestinal and the urogenital tracts, they must possess a greater number of iron-siderophore transporters for cellular respiration and survival. By contrast, the E24 (ETEC) pathovar possesses the same number of iron transporters and OMRs as the nonpathogenic K12, despite the fact that it occupies the small intestines, as do numerous other enteropathogenic strains. This finding suggests that although increases in iron sequestration increase the survival of the pathovars, they are not essential for all of them.

It has been shown that ABC porters (3.A.1) in general possess greater affinity for their substrates than secondary carriers (TC#2.A). Secondary carriers transporting metal cations typically exhibit Km values of ~ 20-800 uM [121] while primary active transporters may exhibit Km values ≤ 1 uM [122]. Interestingly, the majority of ABC iron transporters were found either exclusively in pathogenic strains or in increased numbers (Table 7). This is congruent with our finding that the UPEC strains and APE

possess the largest numbers of iron-siderophore OMRs. TC subclass 2.A iron uptake secondary carriers appear to be present in invariant numbers in all eight *E. coli* strains although the ABC iron-specific transporters differ greatly among the different serovars (Table 7). This suggests that the low affinity secondary iron carriers serve housekeeping functions outside of the body, where iron concentrations are relatively high, while the ABC transporters function in pathogenesis in the *in vivo* environment, where competition with the host is a significant factor.

EHEC and EPEC pathovars typically operate by attaching and effacing (A/E) mechanisms. This type of invasion includes utilizing a Type III secretion system to inject effector proteins into host cells. Typically, these effectors can lead to disruption of host cell signaling pathways [123]. One effector, EspF, plays an important role in tight junction disruption as well as inhibition of sodium and water uptake [123]. In our study, the EspB/D complex was identified in UMN (UPEC) and O15 (EHEC) and serves to initiate pore formation in the host cell plasma membrane as well as translocation of the EspF effector into the cell. O15 is the prototypical EHEC pathovar that utilizes the A/E mechanism. On the other hand, UMN was not expected to possess an EspB/D system as UPEC strains typically utilize Type I Secretion Systems for adhesion and entry through manipulating the host cell cytoskeleton using effectors that stimulate Rho-family GTPases and kinases [124]. Nevertheless, a complete T3SS was identified in UMN, suggesting that an A/E mechanism of invasion may occur in this pathovar. Further studies will be necessary to test this possibility.

The identification of Haemolysin A (HlyA) exclusively in the UPEC strains, ABU and CFT, supports a previous finding that UPEC strains can utilize haemolysins to inhibit protein kinase B signaling and lead to apoptosis of the host cell [37]. Similarly, the exclusive presence of shiga toxin (Stx) in O15 supports previous findings of EHEC pathogenicity [125]. The Stx pathway may provide an alternative pathway by which the EHEC strain can inhibit chemical signaling and immune responses in the host cell.

Homologs of the clostridial cytotoxin (CCT) form channels and inactivate Rho-type GTPases, leading to manipulation of the host cytoskeleton [42]. Predictably, such a toxin was identified in the UPEC strain ABU, but more surprisingly, it was also found in O15. This may suggest an alternative mechanism by which EHEC strains manipulate the host's cytoskeletal machinery since these pathovars typically utilize the mitochondrial-associated protein [78] to inhibit the host cell control protein 42 (CDC42), which provides regulatory functions in host cell actin dynamic control [126]. Additionally, the homologs identified in ABU and O15 were 85% identical, suggesting that the same mechanism is operative. Further studies can be conducted to confirm these suggestions.

AIEC strains such as APE, have been cited in 36% of Crohn's disease. These pathovars operate through necrosis factors that impair host immune responses, allowing the bacteria to colonize the ileal mucosa and lamina propia [127]. We identified the active subunit of Cytolethal Distending Toxin (CdtB) only in APE. The CdtB active subunit induces DNA double-strand breakage, leading to host cell cycle arrest in the G2/M phase [128]. This is preferable in the AIEC pathotype, presumably because pauses

and irregularities in host cell cycle give the bacteria time to proliferate and colonize host cells. Rapid proliferation of these AIEC cells can then induce secretion of Tumor Necrosis Factor Alpha (TNF-α) that causes inflammation in host cells.

Serratia-type pore-forming toxins (S-PFT) were identified in ABU (UPEC), CFT (UPEC) and O15 (EHEC). These toxins exhibit properties vastly different from those of RTX pore-forming haemolysins, which are exported via two-partner secretion systems and are much larger in size (3000+ aas in length vs. 1023 aas in RTX HlyA) [43]. The presence of two-partner secretion system(s) (TPS) in both ABU and O15 confirmed the possibility of a novel *Serratia*-type Pore-forming Toxin (S-PFT) in both UPEC and EHEC strains (see Table 3). Additional studies can be employed to examine their mechanisms of action in UPEC and EHEC pathogenesis.

Several autotransporter-1 systems were identified in the seven pathovars examined. Examples include the fibronectin binding protein (1.B.12.1.3), a tracheal colonization factor (1.B.12.2.2), a firmicute homolog (1.B.12.2.4), and a *Salmonella* BigA homolog (1.B.12.5.5) (see Table 5). Of these autotransporters, fibronectin binding protein (1.B.12.1.3) is known to assist in host cell adhesion in *S. typhimurium* [129], providing evidence that a similar mechanism of adhesion may be utilized by UPEC, AIEC, ETEC and EAEC strains for host cell targeting. Interestingly, homologs of the *B. pertussis* tracheal colonization factor were identified in all pathovars in our study except for O15 (EHEC). Cilia goblet and basal cells primarily make up the pseudostratified squamous epithelium in both mice and humans [130]. The identification of homologs of

the tracheal colonization factor in *E. coli* pathovars may provide clues as to how such strains target the cilia-rich epithelia of both the small and large intestines in host organisms. Overall, ABU contained the most autotransporters, but other pathovars contained at least eight non-orthologous homologs. Compared to K12, which possesses only five autotransporter-1 proteins, the pathogenic strains contained greater numbers with several homologs being absent in K12. This suggests that these additional outer membrane protein secretion systems may assist pathovars in adhering to host cells.

Several autotransporter-2 systems were found exclusively in the pathovars, particularly in APE (AIEC), UMN (UPEC) and O15 (EHEC). Homologs of the membrane-anchored haemagglutinin (1.B.40.1.2) were identified in UMN and O15. These proteins may play roles in bacterium-host cell adhesion and can be responsible for agglutination properties in host erythrocytes as witnessed in other pathovars included in this study [19]. Trimeric adhesins such as Cha (1.B.40.1.5) and UpaG (1.B.40.2.3) were also identified exclusively in pathovars, and they are known to function in adhesion to the host extracellular matrix, notably fibronectin, collagen and intimin [131]. YadB homologs (1.B.40.1.3) function by a similar mechanism, and one was identified in O15, suggesting that it may function specifically in EHEC strains. In K12, no autotransporter-2 complex was identified, suggesting a lack of these adhesion mechanisms that characterize pathogens. Future studies can consider how *E. coli* autotransporter-2 systems mediate certain host-cell adhesion mechanisms.

Although intimins/invasins (autotransporter-3) homologs were identified in all *E. coli* strains, two or more such transporters were observed in all pathovars except 559 (EAEC). In previous studies, EHEC and EPEC strains had been shown to use invasins/intimins to target translocated intimin receptors (Tir) and β-1 integrins on host extracellular matrices for adhesion purposes followed by stimulation of actin rearrangements via the N-Wasp pathway ([132]; [133]). UPEC strains, however, utilize type I pili to bind β-1 integrin on host cell extracellular matrices to initiate actin rearrangements [124]. The presence of invasins/intimins to target β-1 integrins reveals a potential novel mechanism of cell-adhesion and actin rearrangement utilized by these UPEC pathovars. However, further studies are needed to confirm such a proposal.

The MdtP outer membrane factor (OMF; TC# 1.B.17.3.9) functions with the MdtO aromatic acid/drug exporter (2.A.85.6.1) and the auxiliary MFP protein, MdtN (8.A.1.1.3). This OMF was found in all pathovars but not in K12. This may indicate a multi-drug resistance mechanism found exclusively in the pathogenic strains that contributes to their resistance to antibiotics.

Fimbrial usher proteins (FUPs) serve important functions in the assembly of fimbriae [76]. As expected, these proteins were found ubiquitously in all strains including K12. However, greater numbers of genes for types π, γ and α fimbrial usher proteins were observed in all pathovars compared to K12, presumably contributing to pathogenesis by promoting bacterium-host cell adhesion [79].

Two Secretins (1.B.22) were identified exclusively in pathovars. These proteins include the YscC secretin, identified in UMN and O15, and homologs of the phage Gene IV product, which exports phage DNA in certain strains and was identified exclusively in E24 (ETEC). However, the exact functions of these secretins have not been elucidated. Further studies are needed to assess their functions and involvement in *E. coli* pathogenesis, specifically in ETEC, UPEC and EHEC strains.

Potentially complete VasA-L T6SSs were identified in multiple pathovars, including 559 (EAEC), APE (AIEC), E24 (ETEC), UMN (UPEC), and O15 (EHEC). Previous studies have shown that protein products of the *icmF*, *hlc* and *clpV* genes play essential roles in biofilm assembly and general motility in APEC pathovars [134]. In other Gram-negative bacteria, full T6SS proteins provide other functions as well; in recent studies, the T6SS present in *P. aeruginosa* was shown to engage in "T6SS dueling" with surrounding bacteria in competition for survival in the microbiome [135]. Integral components of the T6SS are the T4 phage-like tail, Hcp, and a VgrG "cap" which allows the possessor to infiltrate other bacterial cells [136]. By this mechanism, *P. aeruginosa*, *V. cholerae* and *A. balyli* engage in T6SS battles for occupancy within the microbiome [135].

With the discovery of potentially complete T6SSs in five of the *E. coli* pathovars considered in this study, we speculate that *E. coli* cells may engage in similar mechanisms of T6SS competition within the host microbiome. Having a complete T6SS implies the possession of a novel survival mechanism. It is possible that pathovars of *E.*

coli can utilize T6SSs both for offense and for defense in the host microbiome. Such a mechanism may prove valuable during invasion of the colon, where the microbiome is densely packed with other types of facultative/obligate anaerobes competing for limited metabolic resources.

Adherently invasive pathovars of *E. coli* (AIEC) are known to possess mechanisms for survival within and colonization of host cells [127]. In Crohn's disease, for example, AIEC bind to carcinoembryonic antigen-related cell adhesion molecule 6 (CEACAM6) in host epithelial cells, stimulating increased expression of interferon-γ and tumor necrosis factors [137]. This allows AIEC to alter the cytoskeletal structures of host cells, thereby leading to infiltration and colonization. Similar mechanisms are employed to colonize macrophages [127]. APE, the only AIEC examined in our study, utilizes such mechanisms to colonize host cells.

Uropathogenic *E. coli* (UPEC) strains can colonize transitional cells such as macrophages. Utilizing type 1 fimbriae, these cells bind to α3β1 integrin on host cells that stimulates actin rearrangement via Rac protein signaling [124]. Cytoskeletal rearrangements facilitate engulfment of the bacterial cell. Over time, intracellular bacterial communities form inside the infected host cell through mass colonization [124]. The intracellular bacteremia pathogen ABU and the pyelonephritis pathogen CFT, both included in our study, are UPEC pathogens.

An interesting pattern revealed by our results was the existence of different sets of nutrient uptake transporters, depending on the normal organismal site of proliferation. Intracellular versus extracellular organisms were distinguishable on the basis of the nature of the substrates transported. In general, the intracellular pathovar-specific strains take up metabolites expected to be found inside host cells while the extracellular pathovars transport substrates that are normally found in abundance outside of host cells. Examples of this can be seen in transporters for glycolytic and Kreb cycle intermediates found only in intracellular pathogens as well as transporters for oligosaccharides found only in extracellular pathogens (Table 9). These facts indicate that these organisms have acquired transport systems through the frequent gain and loss of the encoding genes probably via horizontal transfer, to take advantage of nutrients available in their specific environments. Since intracellularity may be a relatively recent development, this suggests that the gain and loss of genes encoding transporters may have occurred with high frequency in response to immediate need. It seems probable that such characteristics (e.g., predilection for certain types of nutrient transporters) can be used for diagnostic purposes in the future as a result of the genomics revolution.

In screening the data we noticed that each strain examined has characteristics that distinguish it from all others. Particularly noteworthy was strain O15, which had numerous characteristics lacking in the other seven strains (section 3.4). One such example includes extra two-partner secretion systems as well as a unique AT-2 adhesin and intimin. More striking was the occurance of large numbers of OmpA-like Ail/Lom family proteins that probably serve multiple virulence functions, including protection

against complement as has been documented in *Yersinia* species [115]. Additionally, this strain possessed numerous holins, all homologous to 1.E.1.1.3. The significance of this last observation, with respect to virulence, is unclear, but none of the other strains examined have this characteristic. It will be interesting to determine if these unique characteristics are causally related to virulence or are purely coincidental.

REFERENCES

1. Peterson JH, Tian P, Ieva R, Dautin N, Bernstein HD: **Secretion of a bacterial virulence factor is driven by the folding of a C-terminal segment**. *Proc Natl Acad Sci U S A* 2010, **107**(41):17739-17744.

2. Williams EJ, Embleton ND, Bythell M, Ward Platt MP, Berrington JE: **The changing profile of infant mortality from bacterial, viral and fungal infection over two decades**. *Acta Paediatr* 2013, **102**(10):999-1004.

3. Zhang L, Foxman B, Manning SD, Tallman P, Marrs CF: **Molecular epidemiologic approaches to urinary tract infection gene discovery in uropathogenic Escherichia coli**. *Infect Immun* 2000, **68**(4):2009-2015.

4. Qadri F, Svennerholm AM, Faruque AS, Sack RB: **Enterotoxigenic Escherichia coli in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention**. *Clin Microbiol Rev* 2005, **18**(3):465-483.

5. Cohen MB, Giannella RA: **Hemorrhagic colitis associated with Escherichia coli O157:H7**. *Adv Intern Med* 1992, **37**:173-195.

6. Kaper JB, Nataro JP, Mobley HL: **Pathogenic Escherichia coli**. *Nat Rev Microbiol* 2004, **2**(2):123-140.

7. Kokesova A, Frolova L, Kverka M, Sokol D, Rossmann P, Bartova J, Tlaskalova-Hogenova H: **Oral administration of probiotic bacteria (E. coli Nissle, E. coli O83, Lactobacillus casei) influences the severity of dextran sodium sulfate-induced colitis in BALB/c mice**. *Folia Microbiol (Praha)* 2006, **51**(5):478-484.

8. Huebner C, Ding Y, Petermann I, Knapp C, Ferguson LR: **The probiotic Escherichia coli Nissle 1917 reduces pathogen invasion and modulates cytokine expression in Caco-2 cells infected with Crohn's disease-associated E. coli LF82**. *Appl Environ Microbiol* 2011, **77**(7):2541-2544.

9. Jonkers D, Penders J, Masclee A, Pierik M: **Probiotics in the management of inflammatory bowel disease: a systematic review of intervention studies in adult patients**. *Drugs* 2012, **72**(6):803-823.

10. Hancock V, Vejborg RM, Klemm P: **Functional genomics of probiotic Escherichia coli Nissle 1917 and 83972, and UPEC strain CFT073: comparison of transcriptomes, growth and biofilm formation**. *Mol Genet Genomics* 2010, **284**(6):437-454.

11. Bhavsar AP, Guttman JA, Finlay BB: **Manipulation of host-cell pathways by bacterial pathogens**. *Nature* 2007, **449**(7164):827-834.

12. Schroeder GN, Hilbi H: **Molecular pathogenesis of Shigella spp.: controlling host cell signaling, invasion, and death by type III secretion**. *Clin Microbiol Rev* 2008, **21**(1):134-156.

13. McDaniel TK, Jarvis KG, Donnenberg MS, Kaper JB: **A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens**. *Proc Natl Acad Sci U S A* 1995, **92**(5):1664-1668.

14. Sokurenko EV, Gomulkiewicz R, Dykhuizen DE: **Source-sink dynamics of virulence evolution**. *Nat Rev Microbiol* 2006, **4**(7):548-555.

15. Levin BR, Bull JJ: **Short-sighted evolution and the virulence of pathogenic microorganisms**. *Trends Microbiol* 1994, **2**(3):76-81.

16. Croxen MA, Finlay BB: **Molecular mechanisms of Escherichia coli pathogenicity**. *Nat Rev Microbiol* 2010, **8**(1):26-38.

17. Ochman H, Selander RK: **Standard reference strains of Escherichia coli from natural populations**. *J Bacteriol* 1984, **157**(2):690-693.

18. Reddy VS, Saier MH, Jr.: **BioV Suite--a collection of programs for the study of transport protein evolution**. *Febs J* 2012, **279**(11):2036-2046.

19. Mancini J, Weckselblatt B, Chung YK, Durante JC, Andelman S, Glaubman J, Dorff JD, Bhargava S, Lijek RS, Unger KP *et al*: **The heat-resistant agglutinin family includes a novel adhesin from enteroaggregative Escherichia coli strain 60A**. *J Bacteriol* 2011, **193**(18):4813-4820.

20. Ikeda M, Arai M, Lao DM, Shimizu T: **Transmembrane topology prediction methods: a re-assessment and improvement by a consensus method using a dataset of experimentally-characterized transmembrane topologies**. *In Silico Biol* 2002, **2**(1):19-33.

21. Zhai Y, Saier MH, Jr.: **A web-based program (WHAT) for the simultaneous prediction of hydropathy, amphipathicity, secondary structure and transmembrane topology for a single protein sequence**. *J Mol Microbiol Biotechnol* 2001, **3**(4):501-502.

22. Reddy VS, Shlykov MA, Castillo R, Sun EI, Saier MH, Jr.: **The major facilitator superfamily (MFS) revisited**. *Febs J* 2012, **279**(11):2022-2035.

23. Castillo R, Saier MH: **Functional Promiscuity of Homologues of the Bacterial ArsA ATPases**. *Int J Microbiol* 2010, **2010**:187373.

24. Saier MH, Jr., Yen MR, Noto K, Tamang DG, Elkan C: **The Transporter Classification Database: recent advances**. *Nucleic Acids Res* 2009, **37**(Database issue):D274-278.

25. Schwarzer C, Fu Z, Patanwala M, Hum L, Lopez-Guzman M, Illek B, Kong W, Lynch SV, Machen TE: **Pseudomonas aeruginosa biofilm-associated homoserine lactone C12 rapidly activates apoptosis in airway epithelia**. *Cell Microbiol* 2012, **14**(5):698-709.

26. Hong SH, Wang X, Wood TK: **Controlling biofilm formation, prophage excision and cell death by rewiring global regulator H-NS of Escherichia coli**. *Microb Biotechnol* 2010, **3**(3):344-356.

27. Durkee JW, Jr., Antich PP, Lee CE: **Exact solutions to the multiregion time-dependent bioheat equation. I: Solution development**. *Phys Med Biol* 1990, **35**(7):847-867.

28. Barabote RD, Saier MH, Jr.: **Comparative genomic analyses of the bacterial phosphotransferase system**. *Microbiol Mol Biol Rev* 2005, **69**(4):608-634.

29. Walther TH, Gottselig C, Grage SL, Wolf M, Vargiu AV, Klein MJ, Vollmer S, Prock S, Hartmann M, Afonin S *et al*: **Folding and self-assembly of the TatA translocation pore based on a charge zipper mechanism**. *Cell* 2013, **152**(1-2):316-326.

30. Davis JK: **Combining polysaccharide biosynthesis and transport in a single enzyme: dual-function cell wall glycan synthases**. *Front Plant Sci* 2012, **3**:138.

31. Hubbard C, McNamara JT, Azumaya C, Patel MS, Zimmer J: **The hyaluronan synthase catalyzes the synthesis and membrane translocation of hyaluronan**. *J Mol Biol* 2012, **418**(1-2):21-31.

32. DeAngelis PL: **Microbial glycosaminoglycan glycosyltransferases**. *Glycobiology* 2002, **12**(1):9R-16R.

33. Denoncin K, Collet JF: **Disulfide bond formation in the bacterial periplasm: major achievements and challenges ahead**. *Antioxid Redox Signal* 2013, **19**(1):63-71.

34. Tzokov SB, Wyborn NR, Stillman TJ, Jamieson S, Czudnochowski N, Artymiuk PJ, Green J, Bullough PA: **Structure of the hemolysin E (HlyE, ClyA, and SheA) channel in its membrane-bound form**. *J Biol Chem* 2006, **281**(32):23042-23049.

35. Hunt S, Moir AJ, Tzokov S, Bullough PA, Artymiuk PJ, Green J: **The formation and structure of Escherichia coli K-12 haemolysin E pores**. *Microbiology* 2008, **154**(Pt 2):633-642.

36. Davies RL, Whittam TS, Selander RK: **Sequence diversity and molecular evolution of the leukotoxin (lktA) gene in bovine and ovine strains of Mannheimia (Pasteurella) haemolytica**. *J Bacteriol* 2001, **183**(4):1394-1404.

37. Wiles TJ, Dhakal BK, Eto DS, Mulvey MA: **Inactivation of host Akt/protein kinase B signaling by bacterial pore-forming toxins**. *Mol Biol Cell* 2008, **19**(4):1427-1438.

38. Iizumi Y, Sagara H, Kabe Y, Azuma M, Kume K, Ogawa M, Nagai T, Gillespie PG, Sasakawa C, Handa H: **The enteropathogenic E. coli effector EspB facilitates microvillus effacing and antiphagocytosis by inhibiting myosin function**. *Cell Host Microbe* 2007, **2**(6):383-392.

39. Aguilera L, Ferreira E, Gimenez R, Fernandez FJ, Taules M, Aguilar J, Vega MC, Badia J, Baldoma L: **Secretion of the housekeeping protein glyceraldehyde-3-phosphate dehydrogenase by the LEE-encoded type III secretion system in enteropathogenic Escherichia coli**. *Int J Biochem Cell Biol* 2012, **44**(6):955-962.

40. Nakao H, Takeda T: **Escherichia coli Shiga toxin**. *J Nat Toxins* 2000, **9**(3):299-313.

41. Pruitt RN, Chambers MG, Ng KK, Ohi MD, Lacy DB: **Structural organization of the functional domains of Clostridium difficile toxins A and B**. *Proc Natl Acad Sci U S A* 2010, **107**(30):13467-13472.

42. Oswald E, Sugai M, Labigne A, Wu HC, Fiorentini C, Boquet P, O'Brien AD: **Cytotoxic necrotizing factor type 2 produced by virulent Escherichia coli modifies the small GTP-binding proteins Rho involved in assembly of actin stress fibers**. *Proc Natl Acad Sci U S A* 1994, **91**(9):3814-3818.

43. Hertle R: **The family of Serratia type pore forming toxins**. *Curr Protein Pept Sci* 2005, **6**(4):313-325.

44. Banerjee J, Singh J, Joshi MC, Ghosh S, Banerjee N: **The cytotoxic fimbrial structural subunit of Xenorhabdus nematophila is a pore-forming toxin**. *J Bacteriol* 2006, **188**(22):7957-7962.

45. Eshraghi A, Maldonado-Arocho FJ, Gargi A, Cardwell MM, Prouty MG, Blanke SR, Bradley KA: **Cytolethal distending toxin family members are**

differentially affected by alterations in host glycans and membrane cholesterol. *J Biol Chem* 2010, **285**(24):18199-18207.

46. Elwell CA, Dreyfus LA: **DNase I homologous residues in CdtB are critical for cytolethal distending toxin-mediated cell cycle arrest**. *Mol Microbiol* 2000, **37**(4):952-963.

47. Smith JL, Bayles DO: **The contribution of cytolethal distending toxin to bacterial pathogenesis**. *Crit Rev Microbiol* 2006, **32**(4):227-248.

48. Saier MH, Ma CH, Rodgers L, Tamang DG, Yen MR: **Protein secretion and membrane insertion systems in bacteria and eukaryotic organelles**. *Adv Appl Microbiol* 2008, **65**:141-197.

49. Wang B, Dukarevich M, Sun EI, Yen MR, Saier MH, Jr.: **Membrane porters of ATP-binding cassette transport systems are polyphyletic**. *J Membr Biol* 2009, **231**(1):1-10.

50. Yen MR, Peabody CR, Partovi SM, Zhai Y, Tseng YH, Saier MH: **Protein-translocating outer membrane porins of Gram-negative bacteria**. *Biochim Biophys Acta* 2002, **1562**(1-2):6-31.

51. Paulsen IT, Park JH, Choi PS, Saier MH, Jr.: **A family of gram-negative bacterial outer membrane factors that function in the export of proteins, carbohydrates, drugs and heavy metals from gram-negative bacteria**. *FEMS Microbiol Lett* 1997, **156**(1):1-8.

52. Saier MH, Jr.: **Protein secretion and membrane insertion systems in gram-negative bacteria**. *J Membr Biol* 2006, **214**(2):75-90.

53. Mueller CA, Broz P, Cornelis GR: **The type III secretion system tip complex and translocon**. *Mol Microbiol* 2008, **68**(5):1085-1095.

54. Woestyn S, Allaoui A, Wattiau P, Cornelis GR: **YscN, the putative energizer of the Yersinia Yop secretion machinery**. *J Bacteriol* 1994, **176**(6):1561-1569.

55. Clements A, Young JC, Constantinou N, Frankel G: **Infection strategies of enteric pathogenic Escherichia coli**. *Gut Microbes* 2012, **3**(2):71-87.

56. Diepold A, Wiesand U, Cornelis GR: **The assembly of the export apparatus (YscR,S,T,U,V) of the Yersinia type III secretion apparatus occurs independently of other structural components and involves the formation of an YscV oligomer**. *Mol Microbiol* 2011, **82**(2):502-514.

57. Alvarez-Martinez CE, Christie PJ: **Biological diversity of prokaryotic type IV secretion systems**. *Microbiol Mol Biol Rev* 2009, **73**(4):775-808.

58. Verma A, Burns DL: **Requirements for assembly of PtlH with the pertussis toxin transporter apparatus of Bordetella pertussis**. *Infect Immun* 2007, **75**(5):2297-2306.

59. Pukatzki S, Ma AT, Revel AT, Sturtevant D, Mekalanos JJ: **Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin**. *Proc Natl Acad Sci U S A* 2007, **104**(39):15508-15513.

60. Loveless BJ, Saier MH, Jr.: **A novel family of channel-forming, autotransporting, bacterial virulence factors**. *Mol Membr Biol* 1997, **14**(3):113-123.

61. Lauri A, Castiglioni B, Mariani P: **Comprehensive analysis of Salmonella sequence polymorphisms and development of a LDR-UA assay for the detection and characterization of selected serotypes**. *Appl Microbiol Biotechnol* 2011, **91**(1):189-210.

62. Etcheverria AI, Padola NL: **Shiga toxin-producing Escherichia coli: factors involved in virulence and cattle colonization**. *Virulence* 2013, **4**(5):366-372.

63. Roggenkamp A, Ackermann N, Jacobi CA, Truelzsch K, Hoffmann H, Heesemann J: **Molecular analysis of transport and oligomerization of the Yersinia enterocolitica adhesin YadA**. *J Bacteriol* 2003, **185**(13):3735-3744.

64. Forman S, Wulff CR, Myers-Morales T, Cowan C, Perry RD, Straley SC: **yadBC of Yersinia pestis, a new virulence determinant for bubonic plague**. *Infect Immun* 2008, **76**(2):578-587.

65. Leo JC, Grin I, Linke D: **Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane**. *Philos Trans R Soc Lond B Biol Sci* 2012, **367**(1592):1088-1101.

66. Gal-Mor O, Gibson DL, Baluta D, Vallance BA, Finlay BB: **A novel secretion pathway of Salmonella enterica acts as an antivirulence modulator during salmonellosis**. *PLoS Pathog* 2008, **4**(4):e1000036.

67. Wentzel A, Christmann A, Adams T, Kolmar H: **Display of passenger proteins on the surface of Escherichia coli K-12 by the enterohemorrhagic E. coli intimin EaeA**. *J Bacteriol* 2001, **183**(24):7273-7284.

68. Aoki SK, Pamma R, Hernday AD, Bickham JE, Braaten BA, Low DA: **Contact-dependent inhibition of growth in Escherichia coli**. *Science* 2005, **309**(5738):1245-1248.

69. Harley KT, Saier MH, Jr.: **A novel ubiquitous family of putative efflux transporters**. *J Mol Microbiol Biotechnol* 2000, **2**(2):195-198.

70. Tseng TT, Gratwick KS, Kollman J, Park D, Nies DH, Goffeau A, Saier MH, Jr.: **The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins**. *J Mol Microbiol Biotechnol* 1999, **1**(1):107-125.

71. Dinh T, Paulsen IT, Saier MH, Jr.: **A family of extracytoplasmic proteins that allow transport of large molecules across the outer membranes of gram-negative bacteria**. *J Bacteriol* 1994, **176**(13):3825-3831.

72. Wong KK, Brinkman FS, Benz RS, Hancock RE: **Evaluation of a structural model of Pseudomonas aeruginosa outer membrane protein OprM, an efflux component involved in intrinsic antibiotic resistance**. *J Bacteriol* 2001, **183**(1):367-374.

73. Letoffe S, Delepelaire P, Wandersman C: **Protein secretion in gram-negative bacteria: assembly of the three components of ABC protein-mediated exporters is ordered and promoted by substrate binding**. *Embo J* 1996, **15**(21):5804-5811.

74. Bleuel C, Grosse C, Taudte N, Scherer J, Wesenberg D, Krauss GJ, Nies DH, Grass G: **TolC is involved in enterobactin efflux across the outer membrane of Escherichia coli**. *J Bacteriol* 2005, **187**(19):6701-6707.

75. Sulavik MC, Houseweart C, Cramer C, Jiwani N, Murgolo N, Greene J, DiDomenico B, Shaw KJ, Miller GH, Hare R *et al*: **Antibiotic susceptibility profiles of Escherichia coli strains lacking multidrug efflux pump genes**. *Antimicrob Agents Chemother* 2001, **45**(4):1126-1136.

76. Clegg S, Wilson J, Johnson J: **More than one way to control hair growth: regulatory mechanisms in enterobacteria that affect fimbriae assembled by the chaperone/usher pathway**. *J Bacteriol* 2011, **193**(9):2081-2088.

77. Henderson NS, So SS, Martin C, Kulkarni R, Thanassi DG: **Topology of the outer membrane usher PapC determined by site-directed fluorescence labeling**. *J Biol Chem* 2004, **279**(51):53747-53754.

78. Mapingire OS, Henderson NS, Duret G, Thanassi DG, Delcour AH: **Modulating effects of the plug, helix, and N- and C-terminal domains on channel properties of the PapC usher**. *J Biol Chem* 2009, **284**(52):36324-36333.

79. Nuccio SP, Baumler AJ: **Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek**. *Microbiol Mol Biol Rev* 2007, **71**(4):551-575.

80. Guedin S, Willery E, Tommassen J, Fort E, Drobecq H, Locht C, Jacob-Dubuisson F: **Novel topological features of FhaC, the outer membrane transporter involved in the secretion of the Bordetella pertussis filamentous hemagglutinin**. *J Biol Chem* 2000, **275**(39):30202-30210.

81. Brok R, Van Gelder P, Winterhalter M, Ziese U, Koster AJ, de Cock H, Koster M, Tommassen J, Bitter W: **The C-terminal domain of the Pseudomonas secretin XcpQ forms oligomeric rings with pore activity**. *J Mol Biol* 1999, **294**(5):1169-1179.

82. Korotkov KV, Gonen T, Hol WG: **Secretins: dynamic channels for protein transport across membranes**. *Trends Biochem Sci* 2011, **36**(8):433-443.

83. McLaughlin LS, Haft RJ, Forest KT: **Structural insights into the Type II secretion nanomachine**. *Curr Opin Struct Biol* 2012, **22**(2):208-216.

84. Sun D, Zhang X, Wang L, Prudhomme M, Xie Z, Martin B, Claverys JP: **Transforming DNA uptake gene orthologs do not mediate spontaneous plasmid transformation in Escherichia coli**. *J Bacteriol* 2009, **191**(3):713-719.

85. Jakes KS, Cramer WA: **Border crossings: colicins and transporters**. *Annu Rev Genet* 2012, **46**:209-231.

86. Cousins RJ, Liuzzi JP, Lichten LA: **Mammalian zinc transport, trafficking, and signals**. *J Biol Chem* 2006, **281**(34):24085-24089.

87. Nishida S, Morinaga Y, Obata H, Mizuno T: **Identification of the N-terminal region of TjZNT2, a Zrt/Irt-like protein family metal transporter, as a novel functional region involved in metal ion selectivity**. *Febs J* 2011, **278**(5):851-858.

88. Karlinsey JE, Maguire ME, Becker LA, Crouch ML, Fang FC: **The phage shock protein PspA facilitates divalent metal transport and is required for virulence of Salmonella enterica sv. Typhimurium**. *Mol Microbiol* 2010, **78**(3):669-685.

89. Gaither LA, Eide DJ: **The human ZIP1 transporter mediates zinc uptake in human K562 erythroleukemia cells**. *J Biol Chem* 2001, **276**(25):22258-22264.

90. Techau ME, Valdez-Taubas J, Popoff JF, Francis R, Seaman M, Blackwell JM: **Evolution of differences in transport function in Slc11a family members**. *J Biol Chem* 2007, **282**(49):35646-35656.

91. Marlovits TC, Haase W, Herrmann C, Aller SG, Unger VM: **The membrane protein FeoB contains an intramolecular G protein essential for Fe(II) uptake in bacteria**. *Proc Natl Acad Sci U S A* 2002, **99**(25):16243-16248.

92. Koster S, Wehner M, Herrmann C, Kuhlbrandt W, Yildiz O: **Structure and function of the FeoB G-domain from Methanococcus jannaschii**. *J Mol Biol* 2009, **392**(2):405-419.

93. Ash MR, Guilfoyle A, Clarke RJ, Guss JM, Maher MJ, Jormakka M: **Potassium-activated GTPase reaction in the G Protein-coupled ferrous iron transporter B**. *J Biol Chem* 2010, **285**(19):14594-14602.

94. Deshpande CN, McGrath AP, Font J, Guilfoyle AP, Maher MJ, Jormakka M: **Structure of an atypical FeoB G-domain reveals a putative domain-swapped dimer**. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2013, **69**(Pt 4):399-404.

95. Ash MR, Maher MJ, Guss JM, Jormakka M: **The initiation of GTP hydrolysis by the G-domain of FeoB: insights from a transition-state complex structure**. *PLoS One* 2011, **6**(8):e23355.

96. Mu W, Zhang W, Feng Y, Jiang B, Zhou L: **Recent advances on applications and biotechnological production of D-psicose**. *Appl Microbiol Biotechnol* 2012, **94**(6):1461-1467.

97. Boudebbouze S, Maguin E, Rhimi M: **Bacterial L-arabinose isomerases: industrial application for D-tagatose production**. *Recent Pat DNA Gene Seq* 2011, **5**(3):194-201.

98. Iyer R, Williams C, Miller C: **Arginine-agmatine antiporter in extreme acid resistance in Escherichia coli**. *J Bacteriol* 2003, **185**(22):6556-6561.

99. Mathews S, Rathinasabapathi B, Ma LQ: **Uptake and translocation of arsenite by Pteris vittata L.: effects of glycerol, antimonite and silver**. *Environ Pollut* 2011, **159**(12):3490-3495.

100. Rothe M, Alpert C, Loh G, Blaut M: **Novel insights into E. coli's hexuronate metabolism: KduI facilitates the conversion of galacturonate and**

glucuronate under osmotic stress conditions. *PLoS One* 2013, **8**(2):e56906.

101. King CD, Rios GR, Green MD, Tephly TR: **UDP-glucuronosyltransferases**. *Curr Drug Metab* 2000, **1**(2):143-161.

102. Severi E, Hood DW, Thomas GH: **Sialic acid utilization by bacterial pathogens**. *Microbiology* 2007, **153**(Pt 9):2817-2822.

103. Goller CC, Seed PC: **Revisiting the Escherichia coli polysaccharide capsule as a virulence factor during urinary tract infection: contribution to intracellular biofilm development**. *Virulence* 2010, **1**(4):333-337.

104. Robicsek F, Citron DS, Taylor FH, Sanger PW: **The Treatment of Malignant Hypertension Due to Ischemia of the Solitary Functioning Kidney**. *Angiology* 1963, **14**:377-380.

105. Rudd PM, Wormald MR, Dwek RA: **Sugar-mediated ligand-receptor interactions in the immune system**. *Trends Biotechnol* 2004, **22**(10):524-530.

106. Peter-Katalinic J: **Methods in enzymology: O-glycosylation of proteins**. *Methods Enzymol* 2005, **405**:139-171.

107. Edmands SD, Hughs KS, Lee SY, Meyer SD, Saari E, Yancey PH: **Time-dependent aspects of osmolyte changes in rat kidney, urine, blood and lens with sorbinil and galactose feeding**. *Kidney Int* 1995, **48**(2):344-353.

108. Anderson PM, Sung YC, Fuchs JA: **The cyanase operon and cyanate metabolism**. *FEMS Microbiol Rev* 1990, **7**(3-4):247-252.

109. Sung YC, Fuchs JA: **Identification and characterization of a cyanate permease in Escherichia coli K-12**. *J Bacteriol* 1989, **171**(9):4674-4678.

110. Pedemonte N, Caci E, Sondo E, Caputo A, Rhoden K, Pfeffer U, Di Candia M, Bandettini R, Ravazzolo R, Zegarra-Moran O *et al*: **Thiocyanate transport in resting and IL-4-stimulated human bronchial epithelial cells: role of pendrin and anion channels**. *J Immunol* 2007, **178**(8):5144-5153.

111. Abdulhussein AA, Wallace HM: **Polyamines and membrane transporters**. *Amino Acids* 2013.

112. Andersson AC, Henningsson S, Rosengren E: **Increased formation of diamines and polyamines in the pregnant rat**. *J Physiol* 1978, **285**:311-324.

113. Gatzeva-Topalova PZ, Walton TA, Sousa MC: **Crystal structure of YaeT: conformational flexibility and substrate recognition**. *Structure* 2008, **16**(12):1873-1881.

114. Choi PS, Dawson AJ, Bernstein HD: **Characterization of a novel two-partner secretion system in Escherichia coli O157:H7**. *J Bacteriol* 2007, **189**(9):3452-3461.

115. Kolodziejek AM, Hovde CJ, Minnich SA: **Yersinia pestis Ail: multiple roles of a single protein**. *Front Cell Infect Microbiol* 2012, **2**:103.

116. Neely MN, Friedman DI: **Functional and genetic analysis of regulatory regions of coliphage H-19B: location of shiga-like toxin and lysis genes suggest a role for phage functions in toxin release**. *Mol Microbiol* 1998, **28**(6):1255-1267.

117. Rainnie DJ, Bragg PD: **The effect of iron deficiency on respiration and energy-coupling in Escherichia coli**. *J Gen Microbiol* 1973, **77**(2):339-349.

118. Wang L, Cherayil BJ: **Ironing out the wrinkles in host defense: interactions between iron homeostasis and innate immunity**. *J Innate Immun* 2009, **1**(5):455-464.

119. Semrin G, Fishman DS, Bousvaros A, Zholudev A, Saunders AC, Correia CE, Nemeth E, Grand RJ, Weinstein DA: **Impaired intestinal iron absorption in Crohn's disease correlates with disease activity and markers of inflammation**. *Inflamm Bowel Dis* 2006, **12**(12):1101-1106.

120. Hancock V, Ferrieres L, Klemm P: **The ferric yersiniabactin uptake receptor FyuA is required for efficient biofilm formation by urinary tract infectious Escherichia coli in human urine**. *Microbiology* 2008, **154**(Pt 1):167-175.

121. Cailliatte R, Schikora A, Briat JF, Mari S, Curie C: **High-affinity manganese uptake by the metal transporter NRAMP1 is essential for Arabidopsis growth in low manganese conditions**. *Plant Cell* 2010, **22**(3):904-917.

122. Anderson DS, Adhikari P, Nowalk AJ, Chen CY, Mietzner TA: **The hFbpABC transporter from Haemophilus influenzae functions as a binding-protein-dependent ABC transporter with high specificity and affinity for ferric iron**. *J Bacteriol* 2004, **186**(18):6220-6229.

123. Hodges K, Alto NM, Ramaswamy K, Dudeja PK, Hecht G: **The enteropathogenic Escherichia coli effector protein EspF decreases sodium hydrogen exchanger 3 activity**. *Cell Microbiol* 2008, **10**(8):1735-1745.

124. Eto DS, Jones TA, Sundsbak JL, Mulvey MA: **Integrin-mediated host cell invasion by type 1-piliated uropathogenic Escherichia coli**. *PLoS Pathog* 2007, **3**(7):e100.

125. Gobert AP, Vareille M, Glasser AL, Hindre T, de Sablet T, Martin C: **Shiga toxin produced by enterohemorrhagic Escherichia coli inhibits PI3K/NF-kappaB signaling pathway in globotriaosylceramide-3-negative human intestinal epithelial cells**. *J Immunol* 2007, **178**(12):8168-8174.

126. Huang Z, Sutton SE, Wallenfang AJ, Orchard RC, Wu X, Feng Y, Chai J, Alto NM: **Structural insights into host GTPase isoform selection by a family of bacterial GEF mimics**. *Nat Struct Mol Biol* 2009, **16**(8):853-860.

127. Rolhion N, Darfeuille-Michaud A: **Adherent-invasive Escherichia coli in inflammatory bowel disease**. *Inflamm Bowel Dis* 2007, **13**(10):1277-1283.

128. Thelestam M, Frisan T: **Cytolethal distending toxins**. *Rev Physiol Biochem Pharmacol* 2004, **152**:111-133.

129. Tukel C, Akcelik M, de Jong MF, Simsek O, Tsolis RM, Baumler AJ: **MarT activates expression of the MisL autotransporter protein of Salmonella enterica serotype Typhimurium**. *J Bacteriol* 2007, **189**(10):3922-3926.

130. Kim JK, Vinarsky V, Wain J, Zhao R, Jung K, Choi J, Lam A, Pardo-Saganta A, Breton S, Rajagopal J *et al*: **In vivo imaging of tracheal epithelial cells in mice during airway regeneration**. *Am J Respir Cell Mol Biol* 2012, **47**(6):864-868.

131. Koretke KK, Szczesny P, Gruber M, Lupas AN: **Model structure of the prototypical non-fimbrial adhesin YadA of Yersinia enterocolitica**. *J Struct Biol* 2006, **155**(2):154-161.

132. Swimm AI, Kalman D: **Cytosolic extract induces Tir translocation and pedestals in EPEC-infected red blood cells**. *PLoS Pathog* 2008, **4**(1):e4.

133. Schuller S, Chong Y, Lewin J, Kenny B, Frankel G, Phillips AD: **Tir phosphorylation and Nck/N-WASP recruitment by enteropathogenic and enterohaemorrhagic Escherichia coli during ex vivo colonization of human intestinal mucosa is different to cell culture models**. *Cell Microbiol* 2007, **9**(5):1352-1364.

134. de Pace F, Boldrin de Paiva J, Nakazato G, Lancellotti M, Sircili MP, Guedes Stehling E, Dias da Silveira W, Sperandio V: **Characterization of IcmF of the type VI secretion system in an avian pathogenic Escherichia coli (APEC) strain**. *Microbiology* 2011, **157**(Pt 10):2954-2962.

135. Basler M, Ho BT, Mekalanos JJ: **Tit-for-tat: type VI secretion system counterattack during bacterial cell-cell interactions**. *Cell* 2013, **152**(4):884-894.

136. Pell LG, Kanelis V, Donaldson LW, Howell PL, Davidson AR: **The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system**. *Proc Natl Acad Sci U S A* 2009, **106**(11):4160-4165.

137. Barnich N, Carvalho FA, Glasser AL, Darcha C, Jantscheff P, Allez M, Peeters H, Bommelaer G, Desreumaux P, Colombel JF *et al*: **CEACAM6 acts as a receptor for adherent-invasive E. coli, supporting ileal mucosa colonization in Crohn disease**. *J Clin Invest* 2007, **117**(6):1566-1574.

APPENDIX

Table 1. Seven *E. coli* pathovars, the non-virulent *E. coli* K12 strain, and their basic traits.

Strain	Abbr.	Genome size (Mbp)	# of Proteins	Host Location	Pathological condition	Pathovar Type
55989	559	5.15	4,759	Extracellular	Diarrhoea	EAEC
ABU 83972	ABU	5.13	4,796	Intracellular	Bacteremia	UPEC
APEC O1	APE	5.50	4,853	Intracellular	Respiratory infection (avians)	AIEC
CFT073	CFT	5.23	5,369	Intracellular	Urinary tract infection & pyelonephritis	UPEC
E24377A	E24	5.25	4,991	Extracellular	Enterotoxicity	ETEC
UMN026	UMN	5.36	5,014	Extra/Intra -cellular	Urinary tract infection	UPEC
O157:H7	O15	5.70	5,477	Extracellular	Haemorrhagic colitis	EHEC
K12 MG1655	K12	4.64	4,290	Extracellular	None	Non-virulent

Table 2. Numbers of proteins (even numbered columns) and percent distribution of subclasses by TC number (odd numbered columns) for all eight strains of *E. coli* as well as numbers of transport proteins per strain (bottom) and all strains combined, with and without K12 (last four columns).

Subclass	559	%	CFT	%	ABU	%	APE	%	E24	%	UMN	%	O15	%	K12	%	all	%	All-K12	%
1.A	42	3.4	37	2.9	41	3.2	41	3.3	42	3.5	39	3.1	42	3.47	39	3.5	323	3.3	284	3.3
1.B	94	7.6	103	8.0	108	8.4	91	7.3	89	7.4	95	7.5	96	7.93	67	5.9	743	7.5	676	7.7
1.C	15	1.2	18	1.4	20	1.6	15	1.2	11	0.9	15	1.2	21	1.74	9	0.8	124	1.3	115	1.3
1.E	5	0.4	6	0.5	4	0.3	6	0.5	3	0.2	7	0.6	14	1.16	4	0.4	49	0.5	45	0.5
2.A	318	25.6	325	25.4	319	24.8	315	25.3	309	25.7	319	25.2	299	24.71	311	27.6	2515	25.5	2204	25.2
2.C	5	0.4	5	0.4	4	0.3	4	0.3	3	0.2	4	0.3	3	0.25	4	0.4	32	0.3	28	0.3
3.A	358	28.8	378	29.5	386	30.0	369	29.7	353	29.4	390	30.8	369	30.50	340	30.1	2943	29.8	2603	29.8
3.B	4	0.3	3	0.2	5	0.4	5	0.4	5	0.4	4	0.3	4	0.33	4	0.4	34	0.3	30	0.3
3.D	69	5.5	66	5.2	64	5.0	59	4.7	69	5.7	72	5.7	68	5.62	69	6.1	536	5.4	467	5.3
4.A	56	4.5	62	4.8	62	4.8	61	4.9	55	4.6	61	4.8	51	4.21	49	4.3	457	4.6	408	4.7
4.B	2	0.2	4	0.3	4	0.3	4	0.3	2	0.2	2	0.2	2	0.17	2	0.2	22	0.2	20	0.2
4.C	10	0.8	17	1.3	16	1.2	11	0.9	9	0.7	11	0.9	7	0.58	8	0.7	89	0.9	81	0.9
4.D	9	0.7	10	0.8	12	0.9	9	0.7	8	0.7	7	0.6	9	0.74	8	0.7	72	0.7	64	0.7
5.A	43	3.5	44	3.4	44	3.4	45	3.6	42	3.5	43	3.4	43	3.55	45	4.0	349	3.5	304	3.5
5.B	3	0.2	2	0.2	2	0.2	2	0.2	2	0.2	4	0.3	3	0.25	3	0.3	21	0.2	18	0.2
8.A	37	3.0	45	3.5	43	3.3	40	3.2	39	3.2	42	3.3	37	3.06	33	2.9	316	3.2	283	3.2
8.B	1	0.1	1	0.1	1	0.1	1	0.1	1	0.1	1	0.1	1	0.08	1	0.1	8	0.1	7	0.1
9.A	87	7.0	64	5.0	63	4.9	77	6.2	71	5.9	66	5.2	60	4.96	42	3.7	530	5.4	488	5.6
9.B	86	6.9	91	7.1	89	6.9	88	7.1	89	7.4	85	6.7	81	6.69	90	8.0	699	7.1	609	7.0
total	1244	100	1281	100	1287	100	1243	100	1202	100	1267	100	1210	100	1128	100	9862	100	8734	100

Table 3. Toxins identified in *E. coli* pathovars and K12.

Family	TC ID	Function	559	ABU	APE	CFT	E24	UMN	O15	K12
HlyE	1.C.10.1.1	Pore-forming haemolysin, HlyE	0	1	0	1	0	1	1	1
RTX toxin	1.C.11.1.3	Pore-forming haemolysin, HlyA	0	2	0	1	0	0	0	0
IIITCP	1.C.36.1.1	T3SS: pore-forming complex, EspBD	0		0	0	0	0	2	0
IIITCP	1.C.36.3.1 and 3.2	T3SS: pore-forming complex, SipBD	0	0	0	0	0	2	0	0
IIITCP	1.C.36.5.1	T3SS: pore-forming complex, SseBCD	0	0	0	0	0	0	1	0
Shiga toxin B	1.C.54.1.1	Shiga toxin B, St-B	0	0	0	0	0	0	2	0
Clostridial cytotoxin	1.C.57.3.2 and 3.3	Pore formation; necrosis in host, Cnf	0	1	0	0	0	0	1	0
S-PFT	1.C.75.1.1	Pore-forming haemolysin, ShlA	0	1	0	1	0	0	1	0
Cytotoxic major fimbrial subunit (Mrx-A)	1.C.80.1.1	Adhesive fimbriae (pore formation)	2	4	3	5	2	2	4	2
Cytolethal distending toxin	1.C.98.1.1	Cell cycle arrest/apoptosis, CdtABC	0	0	3	0	0	0	0	0

Table 4. Pathogenicity-related transenvelope protein secretion systems.

Family	TC ID	Function	559	ABU	APE	CFT	E24	UMN	O15	K12
T1SS	3.A.1.105.3 and 105.4	Drug Exporter 1	8	8	8	8	7	8	8	8
	3.A.1.109.4	Protein Exporter 1	0	4	0	4	0	2	2	0
	3.A.1.110.1	Protein Exporter 2	0	3	0	4	0	0	0	0
	3.A.1.113.3	Peptide 3 Exporter	1	1	1	1	1	1	1	1
	3.A.1.134.7	Peptide 7 Exporter	15	14	13	14	13	13	14	13
T2SS GSP	3.A.5.1.1	SEC-SRP complex	7	7	7	7	7	7	7	7
MTB	3.A.15.1.1	Pullulanese secretion system	8	13	14	8	8	7	0	8
	3.A.15.2.1	Pilin secretion/fimbrial assembly system	5	4	4	3	4	5	4	4
	3.A.15.3.1	LSP/ fimbrial assembly system	0	0	0	1	0	1	0	0
T3SS	3.A.6.1.1	Type III Secretion System Complex	4	1	1	1	6	8	13	1
	3.A.6.2.1	Flagellar Protein Export Complex	11	11	11	11	11	17	11	10
T4SS	3.A.7.3.1	Pertussis Toxin Exporter	0	0	0	0	0	4	0	0
	3.A.7.7.1	Trs DNA Transfer Protein Complex	2	2	2	2	2	2	2	2
	3.A.7.11.1	Type IV Beta-Proteobacterial DNA Secretion System	2	1	4	1	2	2	3	1
	3.A.7.13.2	Plasmid pLS20 Conjugation System	1	2	2	1	1	2	1	1
	3.A.7.14.2	Conjugation System Plasmid DNA IntP/TcpA-J	5	4	3	5	5	4	5	5
T6SS	3.A.23.1.1	T6SS VasA-L	19	8	18	7	17	13	13	0
	3.A.23.2.1	T6SS EvpA-P	19	5	6	5	6	2	2	1

Table 5. Outer membrane protein secretion systems.

Family	TC ID	Function	559	ABU	APE	CFT	E24	UMN	O15	K12
FUP	1.B.11.2.1 and 2.2	Type π fimbrial usher protein, PapC and YbgQ	3	3	3	4	3	3	2	1
	1.B.11.3.2, 3.3, 3.5. 3.6 and 3.9	Type γ fimbrial usher protein, MrkC, EcpC, YcbS, YraJ, and FimD	8	8	5	6	6	8	7	6
	1.B.11.4.1	Type α fimbrial usher protein, CfaC	2	1	1	1	2	1	1	1
AT-1	1.B.12.1.1	Adhesin, AidA	1	1	1	1	1	1	1	2
	1.B.12.1.2	Virulence factor, IcsA (VirG)	2	2	0	2	2	0	1	0
	1.B.12.1.3	Fibronectin binding protein, MisL	1	0	1	1	1	2	1	0
	1.B.12.2.1	Pertactin, Ptt	1	0	0	0	1	1	0	1
	1.B.12.2.2	Tracheal colonization factor, TcfA	1	1	1	1	1	1	0	0
	1.B.12.2.3	Resistance to killing proteins, BrkA	0	0	1	0	0	0	0	0
	1.B.12.2.4	Putative adhesin/protease/pertactin	2	3	1	3	0	1	0	0
	1.B.12.4.2 and 4.3	Serine proteases, EspP and Tsh	3	3	1	3	0	1	1	0
	1.B.12.5.5	Virulence factor, BigA	0	1	1	0	0	1	0	0
	1.B.12.8.2	Auto-aggregation adhesin, Flu	1	3	1	3	2	1	1	1
	1.B.12.8.3	Adhesin, TibA	2	2	2	1	2	3	3	1
OMF	1.B.17.1.1 and 1.4	Hemolysin, drug and siderophore export protein, TolC	1	1	1	1	1	2	2	1
	1.B.17.3.9	Multidrug resistance outer membrane export protein, MdtP	1	1	1	1	1	1	1	0
TPS	1.B.20.1.3 and 1.4	Two-partner secretion exporter, CdiB and OptB	0	1	0	0	0	0	2	0
Secretin	1.B.22.1.1	PulD secretin	1	2	2	1	1	1	0	1
	1.B.22.3.3	YscC secretin	0	0	0	0	0	1	2	0
	1.B.22.4.2	HofQ secretin	1	1	1	1	1	1	1	1
	1.B.22.5.1	Phage gene IV protein secretin	0	0	0	0	1	0	0	0
OMIP	1.B.33.1.3	Outer membrane insertion complex	6	6	6	6	6	6	6	6
AT-2	1.B.40.1.2	Haemagglutinin	0	0	0	0	0	1	1	0
	1.B.40.1.3	Adhesin, YadB	0	0	0	0	0	0	1	0
	1.B.40.1.5	Adhesin, Cha	0	0	1	0	0	1	0	0
	1.B.40.2.3	Adhesin, UpaG	1	1	1	1	1	0	1	0
CsgA	1.B.48.1.1	Curli fiber subunit, CsgA	1	1	1	1	1	1	1	1
AT-3	1.B.54.1.1	Intimin/Invasin (putative autotransporter-3), Eae	2	5	4	4	4	5	4	2

Table 6. Outer membrane iron-siderophore receptors.

Family	TC ID	Receptor Specificity	559	ABU	APE	CFT	E24	UMN	O15	K12
OMR	1.B.14.1.1	Ferric coprogen, FhuE	1	1	1	1	1	1	1	1
	1.B.14.1.2	Ferrichrome, FhuA	1	1	1	1	1	1	1	1
	1.B.14.1.3	Ferric enterobactin, IroN	0	1	0	1	0	0	0	0
	1.B.14.1.4	Catecholate, CirA	1	2	2	3	1	1	2	1
	1.B.14.1.9	Catecholate, Flu	1	1	1	1	1	1	1	1
	1.B.14.1.10	Ferroxamine, FoxA	0	1	1	1	0	0	1	0
	1.B.14.1.12	Ferric citrate, FecA	1	1	0	0	1	1	0	1
	1.B.14.1.13	Ferric receptor, CfrA	1	1	0	1	0	1	1	0
	1.B.14.1.15	Ferrichrome, FcuA	0	1	0	1	0	0	0	0
	1.B.14.2.2	Heme, HemA	0	2	1	2	0	1	1	0
	1.B.14.3.1	Cobalamin, BtuB	1	1	1	1	1	1	1	1
	1.B.14.4.1	Cupric ion, OprC	0	1	1	1	0	1	0	0
	1.B.14.7.2	Ferric-yersinabactin, FyuA	2	2	2	1	1	2	1	1
	1.B.14.9.3	Ferrichrome/Aerobactin, IutA	1	1	0	1	0	1	0	0
	2.C.1.1.1	TonB energy-transducing system	1	2	2	2	1	2	1	1
	2.C.1.2.1	TolA energy-transducing system	3	2	2	1	2	2	2	1

Table 7. Inner membrane iron uptake systems.

Family	TC ID	Function	559	ABU	APE	CFT	E24	UMN	O15	K12
CorA	1.A.35.2.1	Divalent cation transport	2	2	2	2	2	2	2	2
ZIP	2.A.5.5.1	Pmf-mediated M^{2+} uptake	1	1	1	1	1	1	1	1
Nramp	2.A.55.3.1	Divalent cation:H^{+} symport	1	1	1	1	1	1	1	1
FeoB	9.A.8.1.1 and 1.9	GTPase-coupled ferrous iron uptake	2	2	2	2	2	2	2	2
FeT	3.A.1.10.3	Ferric iron uptake	1	1	1	1	1	1	1	1
FeCT	3.A.1.14.1	Ferric iron-dicitrate uptake	4	4	0	0	4	4	0	4
	3.A.1.14.2	Ferric iron-enterobactin uptake	4	8	4	8	4	4	4	4
	3.A.1.14.3	Ferric iron-hydroxamate uptake	3	3	3	3	3	3	3	3
	3.A.1.14.8	Ferric-vibrioferrin uptake	0	2	2	2	0	1	1	0
	3.A.1.14.18	Heme uptake	0	3	3	3	0	3	3	0
	3.A.1.14.21	Heme uptake	0	2	0	2	0	0	0	0
MZT	3.A.1.15.7	Mn^{2+} and Fe^{2+} uptake	0	5	4	5	0	4	0	0
BIT	3.A.1.20.1	Iron uptake complex	0	0	0	0	0	0	3	0
Total			16	32	21	30	16	24	19	16

Table 8. Number of OMR and ABC transport proteins associated with iron transport.

Strain	OMR (iron siderophores)	ABC iron siderophore uptake porters
559	9	16
ABU	15	32
APE	9	21
CFT	14	30
E24	6	16
UMN	10	24
O15	9	19
K12	6	16

Table 9. Metabolite transporters unique to intracellular vs. extracellular pathogens.

Family	TCID	Function	Exclusively in 2 of 3 intracellular pathovars	Exclusively in all 3 intracellular pathovars	Exclusively in extracellular pathovars
E-ClC	1.A.13.2.3	Chloride efflux		X	
TEGT	1.A.14.2.2	Acetate transport		X	
FNT	1.A.16.1.2	Formate transport			x
OMF porin	1.B.6.1.2	Solute diffusion		X	
Tsx	1.B.10.1.1	Nucleoside-specific channel		X	
FUP	1.B.11.2.1	Type π fimbrial export		X	
	1.B.11.2.2	Putative fimbrial export		X	
	1.B.11.3.3	Type γ fimbrial export		X	
	1.B.11.3.5	Type γ fimbrial export		X	
AT-1	1.B.12.1.1**	Adhesin export		X	
	1.B.12.2.1	Pertactin (adhesin) export			x
	1.B.12.2.2**	Tracheal colonization factor export		X	
	1.B.12.4.1	Protease export		X	
	1.B.12.8.2**	Adhesin export			x
	1.B.12.8.3	Adhesin/invasin export		X	
OMF	1.B.17.3.4*	Metal cation (Ag^+) export			x
OMA	1.B.18.1.2	Polysaccharide export			x
Gene IV secretin (OMS)	1.B.22.5.1*	Phage export			x
OmpL porin	1.B.35.2.2	Small molecule transport			x
AT-2	1.B.40.1.2*	Haemaglutinin export			x
	1.B.40.1.3*	Adhesin (YadB) export			x
AT-3 (Int/Inv)	1.B.54.1.1*	γ-Intimin export			x
	1.B.54.1.2**	Invasin export		X	
IIITCP (EspB/D)	1.C.36.1.1	Host pore for T3SS			x
ST-B	1.C.54.1.1*	STX pore-forming virulence factor			x
Cnt2	1.C.57.3.3*	Pore-forming virulence factor			x
Lambda holin	1.E.2.1.4*	Holin; autolysis			x

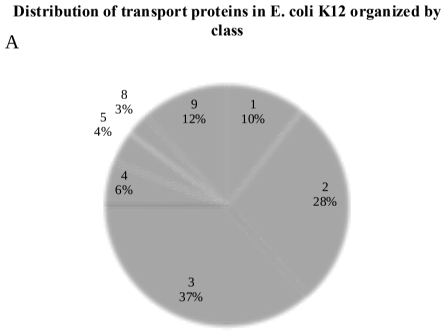
Table 9, Cont.

OMR	1.E.14.1.9	Holin; autolysis		X	
MFS, SP	2.A.1.1.3	Monosaccharide uptake			x
MFS; DHA1	2.A.1.2.7	Drug efflux		x	
MFS; DHA1	2.A.1.2.52	Drug efflux			x
MFS; DHA2	2.A.1.3.18	Efflux of signaling molecules	x		
MFS; OPA	2.A.1.4.2	Glycolytic intermediate uptake	x		
MFS; MHS	2.A.1.6.1*	Tricarboxylate uptake		x	
	2.A.1.6.9	Metabolite uptake		x	
MFS; ACS	2.A.1.14.2	Hexuronate uptake		x	
MFS; AAHS	2.A.1.15.2	Aromatic acid uptake			x
MFS; CP	2.A.1.17.1	Cyanate uptake			x
GPH	2.A.2.1.1	Disaccharide uptake			x
	2.A.2.3.9*	Glycoside uptake			x
APC	2.A.3.1.13	Putrescine (polyamine) uptake			x
	2.A.3.2.7	Arginine uptake; acid homeostasis	x		
	2.A.3.7.2	Glutamate uptake; acid homeostasis			x
	2.A.3.8.17	Amino acid derivative uptake			x
RND	2.A.6.2.19*	Drug export		x	
	2.A.6.2.6	Drug/autoinducer efflux	x		
Betaine/carnitine/choline symporter	2.A.15.2.3	Organoamine uptake			x
CPA1	2.A.36.3.1	Na^+:H^+ antiport		x	
CNT	2.A.41.2.10	Nucleoside uptake			x
DASS	2.A.47.3.2	Citrate uptake; succinate export		x	
	2.A.47.4.2	Arsenite/antimonite export		x	
SulP	2.A.53.3.8	Bicarbonate uptake			x
TRAP-T	2.A.56.1.5**	Rhamnogalacturonide uptake		x	
	2.A.56.1.6	Sialic acid uptake		x	
DcuC	2.A.61.1.2*	Dicarboxylate antiport			x

Table 9 , Cont.

PST	2.A.66.2.10	Lipopolysaccharide precursor export	x		
AbgT	2.A.68.1.1	p-Aminobenzoyl-glutamate uptake			x
TTT	2.A.80.1.1	Tricarboxylate uptake		x	
ABC; CUT2	3.A.1.2.1**	Ribose uptake		x	
	3.A.1.2.8	Sugar uptake			x
	3.A.1.2.11**	Sugar alcohol uptake		x	
	3.A.1.2.13	Autoinducer-2 uptake		x	
ABC; PepT	3.A.1.5.12	Rhamnose oligosaccharide uptake			x
	3.A.1.5.20	Oligopeptide uptake		x	
ABC; LPT	3.A.1.125.2	Lipoprotein export			x
ArsAB	3.A.4.1.1*	Arsenite export			x
NDH	3.D.1.9.1*	H^+/K^+ exchange (Hydrogenase)			x
NFO	3.D.6.1.2	H^+ export; NADH:ferridoxin oxido-reductase			x
PTS; Glc	4.A.1.1.3	Maltose uptake		x	
	4.A.1.1.9	Glucose uptake		x	
	4.A.1.1.10	Glucose uptake			x
PTS; Fru	4.A.2.1.3*	2-O-α-Mannosyl-D glycerate uptake			x
	4.A.2.1.6*	Mannose uptake			x
	4.A.2.1.9	Keto sugar uptake		x	
PTS; Gut	4.A.4.1.2	Sugar alcohol (glucitol) uptake			x
PTS; Gat	4.A.5.1.3	Sugar alcohol (galactitol) uptake			x
PnuC	4.B.1.1.4	Coenzyme precursor (thiamin) uptake		x	
DUF318	9.B.28.1.1*	Unknown			x
YiaW; DUF3302	9.B.32.1.2*	Unknown			x
UIT3	9.B.50.1.2**	Possible dicarboxylate uptake		x	
DUF805	9.B.124.1.1*	Unknown			x
DUF805 protein	9.B.124.1.4	Unknown		x	

A


Distribution of transport proteins in E. coli K12 organized by class
8 3%
9 12%
1 10%
5 4%
4 6%
2 28%
3 37%

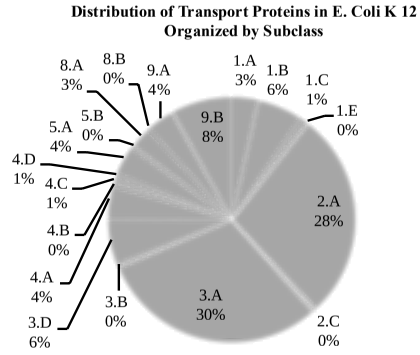

Distribution of Transport Proteins in E. Coli K 12 Organized by Subclass
8.A 3%
8.B 0%
9.A 4%
1.A 3%
1.B 6%
1.C 1%
1.E 0%
5.B 0%
9.B 8%
5.A 4%
4.D 1%
4.C 1%
2.A 28%
4.B 0%
4.A 4%
3.B 0%
3.A 30%
3.D 6%
2.C 0%

Figure 1.

B

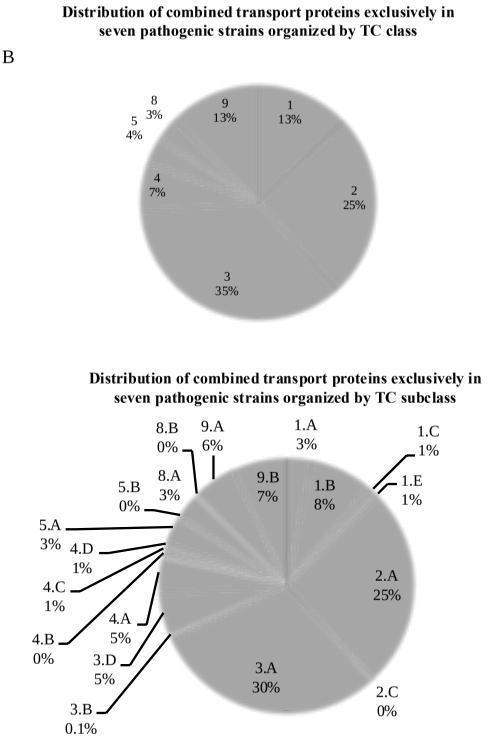


Figure 1, Cont.

C

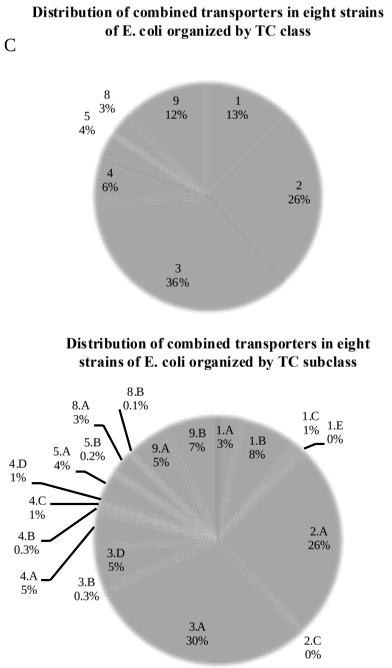


Figure 1, Cont.

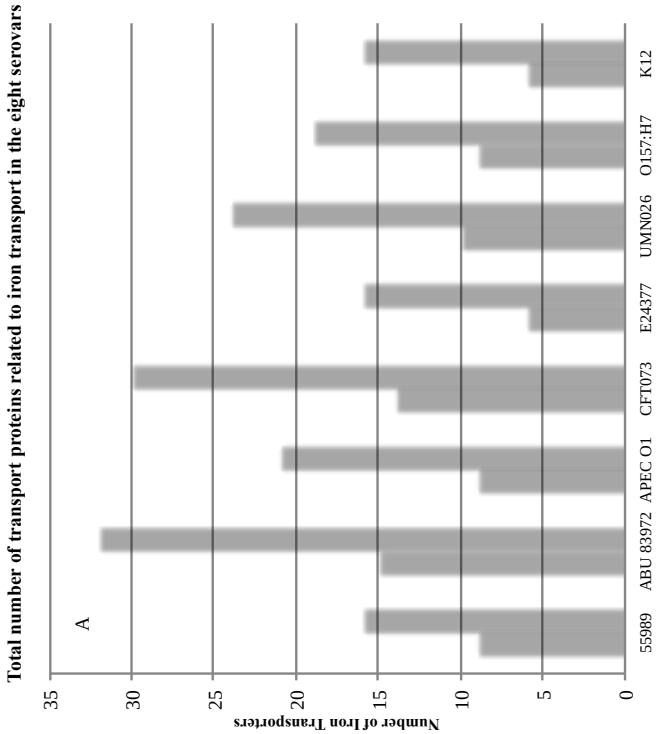

Total number of transport proteins related to iron transport in the eight serovars
A
Number of Iron Transporters
35
30
25
20
15
10
5
0
55989
ABU 83972
APEC O1
CFT073
E24377
UMN026
O157:H7
K12
Strain
OMR siderophores
IM siderophore uptake porters

Figure 2.

B

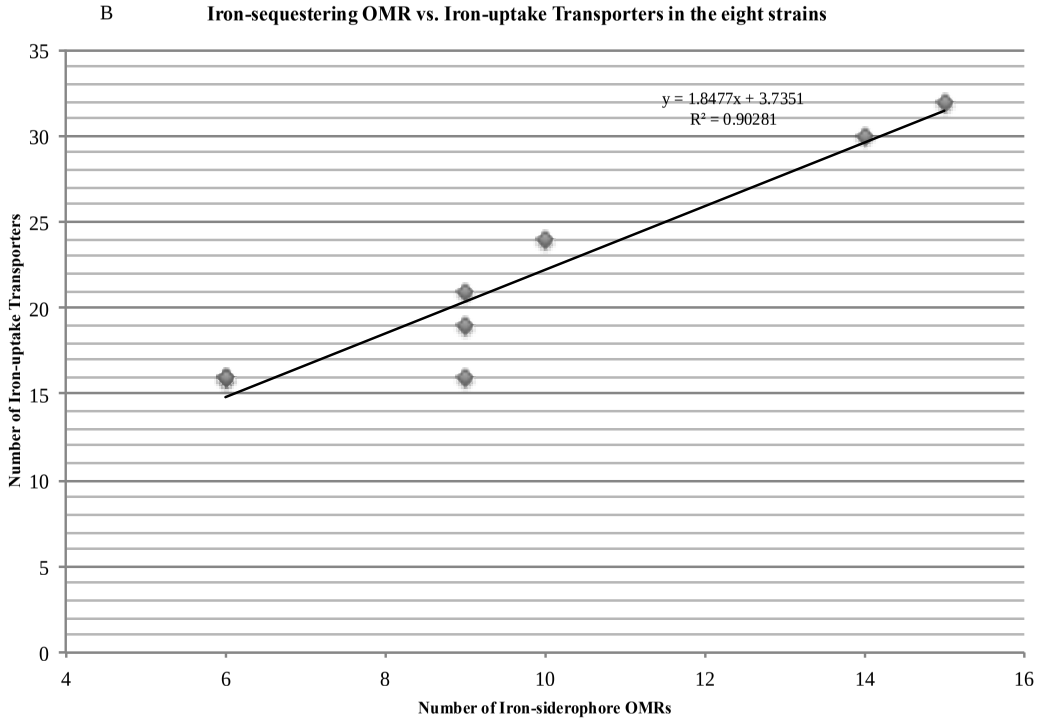


Figure 2, Cont.

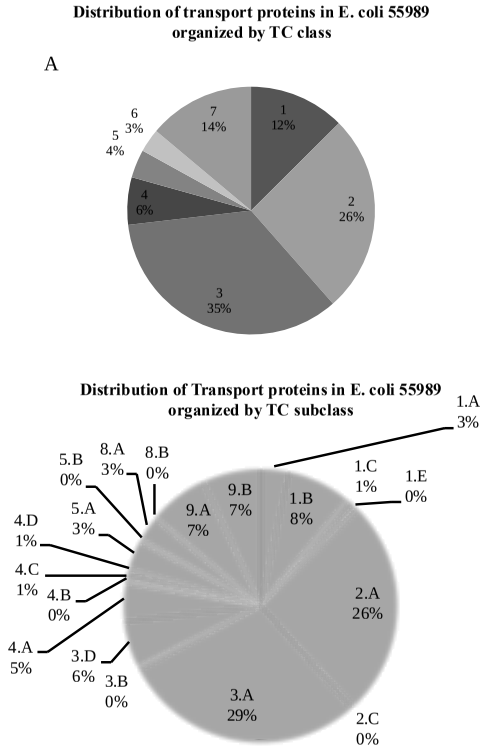

Distribution of transport proteins in E. coli 55989 organized by TC class
A
1 12%
2 26%
3 35%
4 6%
5 4%
6 3%
7 14%
Distribution of Transport proteins in E. coli 55989 organized by TC subclass
1.A 3%
1.B 8%
1.C 1%
1.E 0%
2.A 26%
2.C 0%
3.A 29%
3.B 0%
3.D 6%
4.A 5%
4.B 0%
4.C 1%
4.D 1%
5.A 3%
5.B 0%
8.A 3%
8.B 0%
9.A 7%
9.B 7%

Figure 3.

B

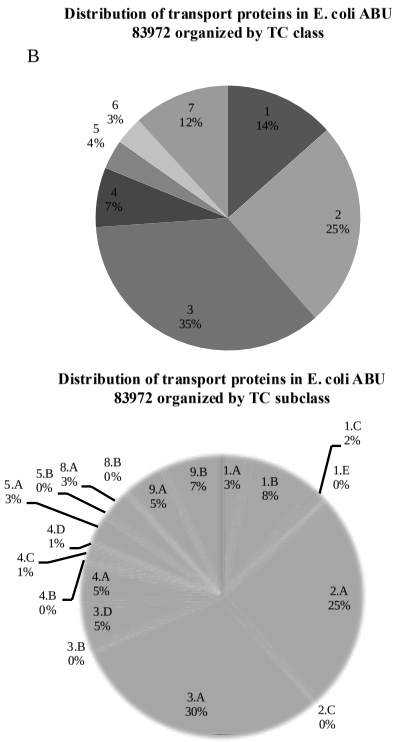


Figure 3, Cont.

C

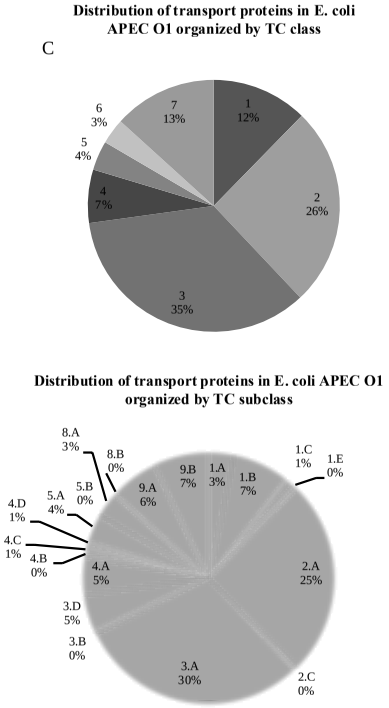


Figure 3, Cont.

D

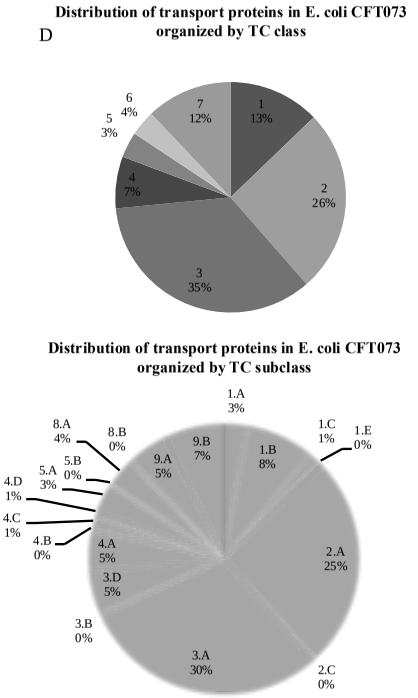


Figure 3, Cont.

E

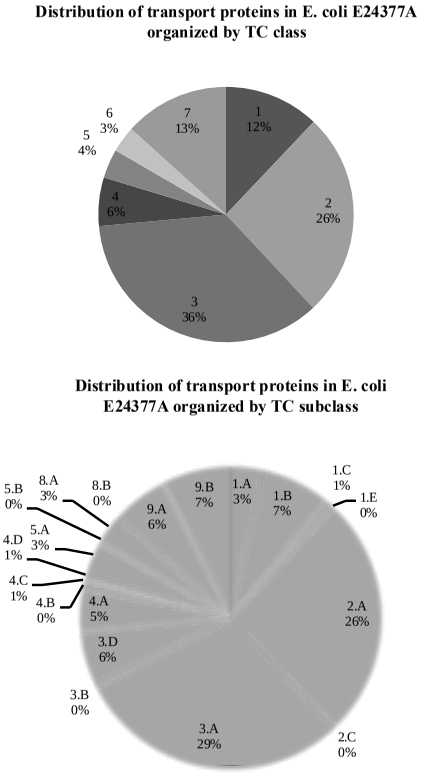


Figure 3, Cont.

F

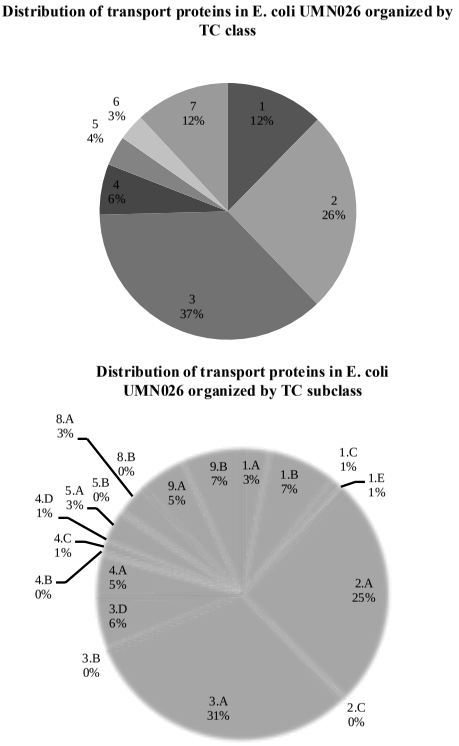


Figure 3, Cont.

G

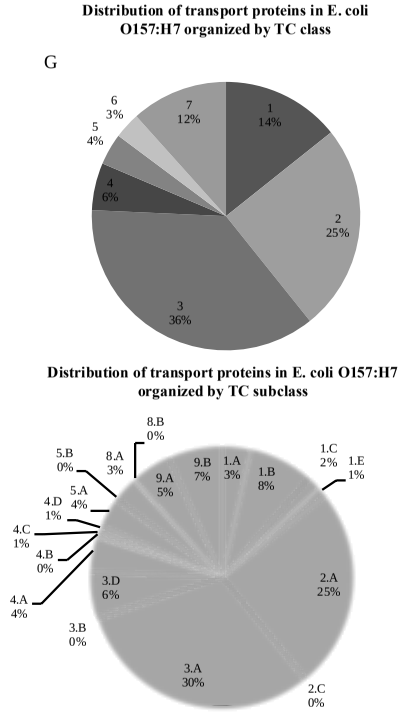


Figure 3, Cont.

H

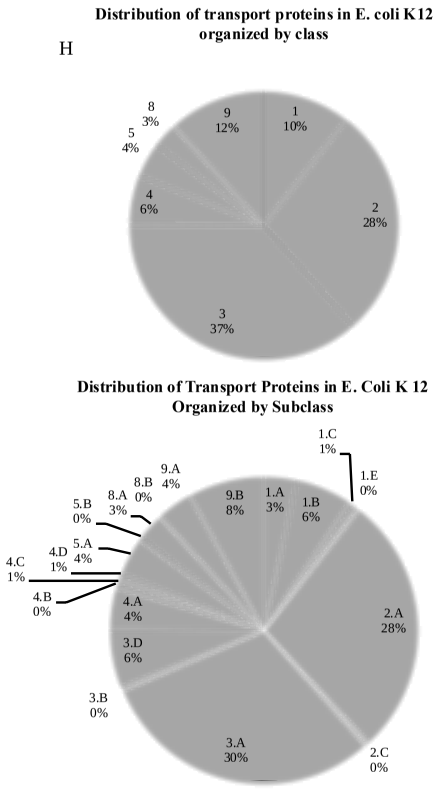


Figure 3, Cont.