



Computational Analysis of Host–Pathogen Protein Interactions between Humans and Different Strains of Enterohemorrhagic *Escherichia coli*

Tungadri Bose^{1,2}, K. V. Venkatesh² and Sharmila S. Mande^{1*}

¹ Bio-Sciences R&D Division, TCS Innovation Labs, Tata Consultancy Services Limited, Pune, India, ² Department of Chemical Engineering, Indian Institute of Technology Bombay, Mumbai, India

OPEN ACCESS

Edited by:

Alfredo G. Torres,
University of Texas Medical Branch,
USA

Reviewed by:

Scott H. Harrison,
North Carolina Agricultural and
Technical State University, USA
Biswapriya Biswas Misra,
Texas Biomedical Research Institute,
USA

*Correspondence:

Sharmila S. Mande
sharmila.mande@tcs.com;
sharmila@atc.tcs.com

Received: 30 November 2016

Accepted: 28 March 2017

Published: 19 April 2017

Citation:

Bose T, Venkatesh KV and Mande SS
(2017) Computational Analysis of
Host–Pathogen Protein Interactions
between Humans and Different Strains
of Enterohemorrhagic *Escherichia coli*.
Front. Cell. Infect. Microbiol. 7:128.
doi: 10.3389/fcimb.2017.00128

Serotype O157:H7, an enterohemorrhagic *Escherichia coli* (EHEC), is known to cause gastrointestinal and systemic illnesses ranging from diarrhea and hemorrhagic colitis to potentially fatal hemolytic uremic syndrome. Specific genetic factors like *ompA*, *nsrR*, and *LEE* genes are known to play roles in EHEC pathogenesis. However, these factors are not specific to EHEC and their presence in several non-pathogenic strains indicates that additional factors are involved in pathogenicity. We propose a comprehensive effort to screen for such potential genetic elements, through investigation of biomolecular interactions between *E. coli* and their host. In this work, an *in silico* investigation of the protein–protein interactions (PPIs) between human cells and four EHEC strains (viz., EDL933, Sakai, EC4115, and TW14359) was performed in order to understand the virulence and host-colonization strategies of these strains. Potential host–pathogen interactions (HPIs) between human cells and the “non-pathogenic” *E. coli* strain MG1655 were also probed to evaluate whether and how the variations in the genomes could translate into altered virulence and host-colonization capabilities of the studied bacterial strains. Results indicate that a small subset of HPIs are unique to the studied pathogens and can be implicated in virulence. This subset of interactions involved *E. coli* proteins like YhdW, ChuT, EivG, and HlyA. These proteins have previously been reported to be involved in bacterial virulence. In addition, clear differences in lineage and clade-specific HPI profiles could be identified. Furthermore, available gene expression profiles of the HPI-proteins were utilized to estimate the proportion of proteins which may be involved in interactions. We hypothesized that a cumulative score of the ratios of bound:unbound proteins (involved in HPIs) would indicate the extent of colonization. Thus, we designed the Host Colonization Index (HCI) measure to determine the host colonization potential of the *E. coli* strains. Pathogenic strains of *E. coli* were observed to have higher HCIs as compared to a non-pathogenic laboratory strain. However, no significant differences among the HCIs of the two pathogenic groups were observed. Overall, our findings are expected to provide additional insights into EHEC pathogenesis and are likely to aid in designing alternate preventive and therapeutic strategies.

Keywords: *Escherichia coli* serotype O157:H7, host–pathogen interaction, protein interaction network, surface and secreted proteins, host colonization index

INTRODUCTION

Escherichia coli O157:H7 has emerged as an important zoonotic pathogen. In recent years, different strains belonging to the O157:H7 serotype of *E. coli* has been reported to be associated with disease outbreaks in U.S., Canada, Japan, and UK (Karmali, 1989; Zhang et al., 2007; Tildesley et al., 2012). It is one of the most prominent classes of enterohemorrhagic *Escherichia coli* (EHEC) and is known to cause gastrointestinal and systemic illnesses (Karmali, 1989; Zhang et al., 2007). In humans, the illness caused by EHEC ranges from diarrhea and hemorrhagic colitis to the more severe and potentially fatal hemolytic uremic syndrome (Karmali, 1989; Karch et al., 2005).

Although the primary habitat of *E. coli* O157:H7 is the ruminant gut, it has also been observed to colonize the human gastrointestinal tract. The infection resulting due to its colonization has been reported to be more severe in humans as compared to that in cattle (Zhang et al., 2010). Consequently, attempts have been made to understand the spatiotemporal nature of immune response in cattle under the influence of *E. coli* O157:H7 infection (Tildesley et al., 2012). The role of bacterial factors, like the type III secretion system, that are essential for colonization and persistence inside the cattle GI tract, have also been extensively studied (Zhang et al., 2010; Tildesley et al., 2012). However, the interplay between the host and the pathogenic factors are poorly understood. A few studies have also reported the effect of *E. coli* O157:H7 colonization on the human immune system (Li et al., 2000; Ho et al., 2013a,b). Studies have suggested the roles of specific bacterial proteins like OmpA (Outer membrane protein A) in *E. coli* (EHEC) pathogenesis (Torres et al., 2006). In addition, the role of NsrR (nitric oxide sensor nitrite-sensitive repressor) in the positive regulation of the *LEE* (Locus of Enterocyte Effacement) genes has also been solicited (Branchu et al., 2014). However, these genetic factors are not specific to EHEC and their presence in genomes of several non-pathogenic strains indicates that additional factors are likely to be involved in manifestation of pathogenicity by the EHEC strains. Thus, a comprehensive effort would be required to screen for (and model interplay between) potential genetic elements which may contribute toward successful propagation and virulence of EHEC strains inside the human body.

In order to understand the mechanism of bacterial association with its host, it is important to study protein–protein interaction (PPI) between the bacterial and the host proteins. Several researchers in the past have explored the avenue of studying “host–pathogen PPI” (HPI) in order to decipher the cross-talks between the host and the pathogen during the course of an infection (Evans et al., 2009; Tyagi et al., 2009; Wuchty, 2011; Zhou et al., 2014). Interestingly, earlier reports have indicated variations in the magnitude of immune responses among cattle calves infected with different strain of *E. coli* (EHEC; Corbishley et al., 2014). A comprehensive study comparing the differences in HPI patterns involving a host and various strains of *E. coli* (EHEC) is unavailable. Minor differences among the pathogenic determinants have also been reported in different strains of *E. coli* (EHEC) which affect the cattle calves (Corbishley et al., 2014). Therefore, inspection of the PPIs among host and different strains of *E. coli* could be helpful in decoding the probable tactics

adopted by the bacteria to counter host defensive strategies. As one would expect, the dynamics of the colonization within the host GI tract (and the subsequent infection) would depend on interactions between the host and the pathogen proteins. In other words, the response of the host (human or cattle) would be governed by not only the genomic differences among bacterial strains, but also the host proteins which interact with the bacterial proteins. It is also well-documented that certain lineage specific variations exist among the genomes of different strains of *E. coli* O157:H7. These include the encoded Shiga toxins, the arginine translocation system, adhesion factors, and other mobile genetic elements (Zhang et al., 2010; Eppinger et al., 2011). Further, *E. coli* O157:H7 strains belonging to such lineages differ in their distribution across human and cattle hosts (Zhang et al., 2010). A comparative analysis of the HPIs would help in understanding probable mechanisms of host (human or cattle) invasion by various strains of *E. coli* (EHEC).

Most of the reported HPI studies have focused on understanding interactions of a specific pathogenic strain with its host (Evans et al., 2009; Tyagi et al., 2009; Wuchty, 2011; Zhou et al., 2014). In order to understand interactions of various strains of *E. coli* with human cells, we have developed a strategy for predicting HPIs between the host and multiple strains of *E. coli* using publicly available information (completely sequenced genomes as well as intra-species PPI data). A comparative analysis involving four EHEC strains of *E. coli* (viz., EDL933, Sakai, EC4115, and TW14359) and a non-pathogenic laboratory strain of *E. coli* (K-12 MG1655) was performed. Notably, while two of the pathogenic strains (EDL933 and Sakai) belonged to lineage I which are equally distributed among humans and cattle, EC4115 and TW14359 belonged to lineage I/II, which are known to share genetic and physiological characteristics intermediate to those of strains from lineages I and II (Zhang et al., 2010; Eppinger et al., 2011). Results of the present study indicate that a small subset of HPIs is unique to the studied pathogens which may contribute toward expression of virulent phenotypes. In addition, clear differences in lineage and clade specific HPI profiles could also be identified. Furthermore, we utilized gene expression profiles of the HPI–proteins to estimate the proportion of proteins which may be involved in the interactions. Based on the ratios of bound protein to total protein concentrations (involved in different HPIs), we designed a metric (called the Host Colonization Index or HCI) to determine the host colonization potential of the *E. coli* strains. We observed that pathogenic strains of *E. coli* had higher HCI as compared to non-pathogenic laboratory strain, thereby, re-emphasizing the vital role of HPIs in bacterial pathogenesis. Overall, our findings are expected to provide additional insights into EHEC pathogenesis and aid in designing alternate preventive and therapeutic strategies.

RESULTS

Protein–Protein Interactions Involving Human Cells and *E. coli*

The inter-species interlogs method (details in Materials and Methods Section) was employed to predict host–pathogen interactions (HPIs) between proteins from humans and those

from the five studied strains of *E. coli*. **Supplementary Table 1** lists the predicted HPIs corresponding to the five studied *E. coli* strains. The interactomes contained 762–783 HPIs and were comprised of 165–171 human proteins and 118–122 *E. coli* proteins. The HPI networks obtained between human and all the studied strains of *E. coli* were observed to have a single major sub-network (comprising of over 125 proteins) and several small sub-networks, each comprising of 2–25 interacting protein.

In order to identify key host as well *E. coli* proteins that are likely to play important roles during pathogenesis, graph properties like degree and betweenness centralities were analyzed from the obtained PPI networks. Degree, a measure of the number of links that any node in the network (in this case, a protein from either host or *E. coli*) shares with the other nodes, is a direct indicator of the importance of that protein in the host–pathogen cross-talks. It was observed that, while the average degree of the networks ranged between 5.2 and 5.4, the degree distribution of some *E. coli* proteins like the L-cystine-binding protein (FliY), UDP-sugar hydrolase (UshA), acyl-CoA thioesterase I (TesA), ferrienterobactin-binding periplasmic protein (FepB), glutathione peroxidase (BtuE), ferric hydroxamate uptake protein D (FhuD), and glutamine-binding periplasmic protein (GlnH), and some human proteins of the ATP-binding cassette like ABCB5, ABCA6, ABCB1, ABCB4, ABCB9, ABCA10, ABCA12, ABCA9, ABCB6, and ABCB8 were appreciably higher (≥ 20). On the other hand, betweenness of a node (protein) was calculated for measuring the number of shortest paths passing through that node. As expected, most of the human and *E. coli* proteins which demonstrated high degree centralities also had high betweenness values. However, it was interesting to note that some human proteins like formyltransferase (ATIC), thymidylate synthetase (TYMS), glutathione S-transferase zeta 1 (GSTZ1), pyruvate dehydrogenase (PDPR), aminomethyltransferase (AMT), cystathione gamma lyase (CTH), glutamic oxaloacetic transaminases 1 and 2 (GOT1 and GOT2) had high betweenness centralities, in spite of low degree distributions. The gene encoding one of these proteins, namely GSTZ1, has been previously reported to have a staggering 20-fold over-expression in *E. coli* infected murine models (Motley et al., 2004). This observation indicates the likely involvement of the identified human proteins in *E. coli* mediated host responses. Furthermore, two *E. coli* proteins, namely, sulfate-binding protein (Sbp) and thiosulfate sulfurtransferase (YnjE) demonstrated high betweenness centralities, in spite of their relatively low degree distributions. Homologs of YnjE have previously been implicated for the virulent properties of *Salmonella enterica* serovar, when studied for systemic infection in a mouse model (Wallrodt et al., 2013). The above observations further indicate a role of these proteins during bacterial colonization in the host gut.

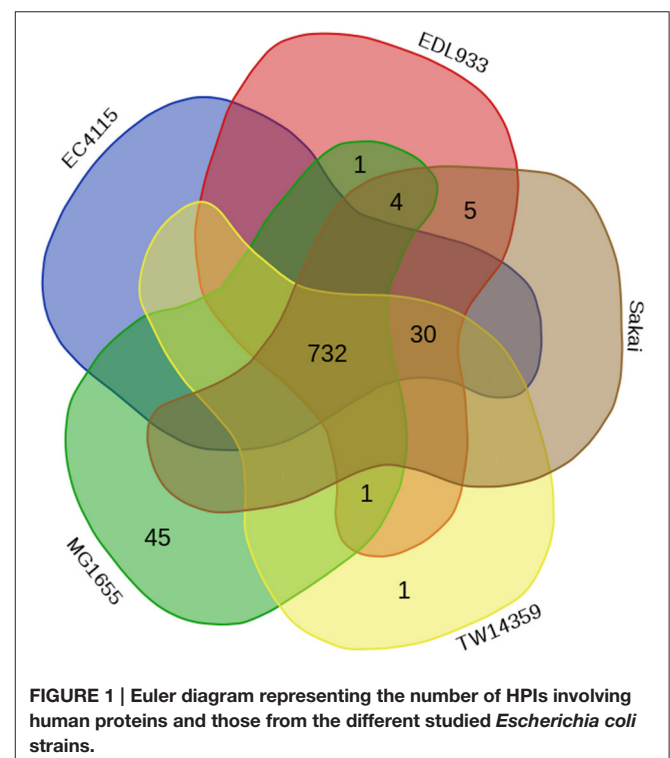
Comparative Analysis of Protein Interactions between Human Cells and Different Strains of *E. coli*

The *E. coli* O157:H7 strains considered in the current study are known to be associated with disease outbreaks of varying

magnitudes in different parts of the world (Zhang et al., 2007; Deng et al., 2011; Eppinger et al., 2011). It may be assumed that the interactome profile of these *E. coli* strains with the human cells would have a significant bearing on the nature of infection caused by these strains. It is therefore important to understand how genomic variations of these strains could affect their (host–pathogen) interactome profiles. Furthermore, whether and how the observed variations in the interactome profile pertaining to these strains translate into altered virulent potentials also remains to be deciphered. In order to evaluate whether proteins from these studied *E. coli* strains interact differently with the human cells, the obtained HPIs between host and different *E. coli* strains were compared. Results of the comparison indicated that although the majority of the interactions between host proteins and those in various *E. coli* strains were common to both pathogenic as well as non-pathogenic strains, 30 interactions were observed to be unique to the four pathogenic strains (**Figure 1**). In contrast, 45 HPIs were noted to be present only in the interactome corresponding to the non-pathogenic *E. coli* strain (**Figure 1**). The above observation suggests distinct differences between the HPI pattern involving human cells pathogenic and non-pathogenic strains of *E. coli*.

Human–*E. coli* Protein Interactions Common to All Studied Strains

The majority (over 90%) of the PPIs were seen to be ubiquitously present in the interactome data involving human cells and the five studied strains of *E. coli*. Literature evidence supporting the biological basis for several of the identified HPIs could be obtained. For example, human nitric oxide synthase (NOS3) was



found to interact with CysP of *E. coli*. Interestingly, NOS3 is involved in the synthesis of peroxynitrite (ONOO⁻), a known agent for innate immunity against enteric pathogens (Henard and Vázquez-Torres, 2011). Further, the level of expression of *cysP* and other related genes (in *E. coli*) have been reported to increase by 4- to 6-fold on exposure to peroxynitrite (McLean et al., 2010). Similarly, the observed interaction between host GGTL1 (gamma-glutamyltransferase light chain) and the pathogenic *gadB* (glutamate decarboxylase) is interesting since *E. coli* GadA and GadB have previously been found to be responsible for the pathogenic ability to withstand the highly acidic environment within the human gut (Ma et al., 2002). In yet another HPI, the predicted interaction between *E. coli* KatG and KatE with superoxide dismutase (SOD2) of the host could find a support from earlier biological findings. While human SOD2 is devised for neutralizing intracellular superoxide (O₂⁻) stress by producing hydrogen peroxide (H₂O₂), the bacterial KatG and KatE enzymes function by detoxifying the H₂O₂-induced stress in *E. coli* (Jung and Kim, 2003).

Several predicted interactions between human and *E. coli* proteins were found to involve bacterial chaperones and other proteins which aid the bacteria in protein folding. This finding is significant since earlier reports have indicated the potency of bacterial chaperones to act as virulence factors, influence host cell–cell signaling, and even promote apoptosis (Henderson et al., 2006). For instance, *E. coli* PpiA was observed to interfere with the functioning of several human proteins including FK506 binding protein 2 (FKBP2), tyrosine phosphatase (PRL), calnexin (CANX), squamous cell carcinoma antigen (SART3), basigin (BSG), transcriptional repressor protein YY1, and prolactin receptor (PRLR). While most of these human proteins were found to be associated with the endoplasmic reticulum, the *E. coli* PpiA, a chaperone, is known to play a vital role in protein folding and is essential for bacterial growth (Justice et al., 2005). DsbA, yet another bacterial protein which is involved in periplasmic disulfide bond formation, was observed to interact with the human intestinal-type alkaline phosphatase (ALPI). ALPI was further observed to interact with key bacterial proteins like SecA (essential for protein export), FolB and FloE (involved in folate biosynthesis), CreC (member of two-component signal transduction machinery), and GlgA (glycogen synthase). Interestingly, it has earlier been proposed that the intestinal alkaline phosphatases have the ability to detoxify (bacterial) lipopolysaccharide and hence, prevent bacterial invasion across the gut mucosal barrier (Goldberg et al., 2008). The identified PPIs involving human ALPI demonstrates that ALPI probably interferes with a larger array of previously unreported bacterial pathways and is important for host defense against an invading *E. coli*.

In the predicted HPI models, bacterial proteins like FhuD and FhuA were observed to interact with various members of the human ABC transporter family like ABCC6, ABCC8, and ABCG8 in all strains of *E. coli*. Bacterial FhuD and FhuA are components of the *fhu* operon which are known to be essential for utilization of ferric siderophores, thereby aiding the pathogen in iron acquisition (Mademidis and Köster, 1998). Given the low availability of iron/heme *in vivo*, the interaction of FhuD and

FhuA with the human ABC transporters may be crucial for the survival of the pathogen.

In spite of the subtle differences in the HPI profiles, the majority of interactions were observed to be commonly present in the interactome data corresponding to the different studied strains of *E. coli*. While some of these interactions appeared to be involved in the mechanism by which the *E. coli* intrude into the host, others were representations of the host's immune response to counter an invading pathogen. This result indicates that the mechanism by which *E. coli* attaches to the host epithelium and subsequently colonizes the host gut is largely ubiquitous.

Human–*E. coli* Protein Interactions Common to All Pathogenic Studied Strains

Although most of the PPIs were common to the interactome data involving both pathogenic and non-pathogenic strains of *E. coli*, a small subset of HPIs (30 interactions) were specifically observed to be associated with the pathogenic strains (Figure 1). Such HPIs include *E. coli* EivG (type III secretion apparatus protein), ChuT (associated with heme sequester and transporter), YhdW (membrane component of amino acid ABC transporter), HlyA (associated with biofilm formation), and a putative outer membrane exporter protein (herein referred as pOMP1). These proteins were seen to primarily interact with various human proteins belonging to the ABC transporter family and the V-type proton ATPase subunit B (ATP6V1B2). ATP6V1B2 which plays a crucial role in phagosomal maturation was found to interact with FlgC (flagellar basal body protein) and EivG (type III secretion apparatus protein) in all strains of *E. coli* O157:H7, but was found to be absent in the interactome data pertaining to MG1655 (the non-pathogenic *E. coli* strain used in this study). Considering the vital role of type III secretion system in the virulence of several gram-negative bacteria, it could be assumed that this HPI may have a significant effect of the manifestation of pathogenesis by the *E. coli* O157:H7 strains.

The host protein ABCA12 was predicted to interact with ChuT and YhdW in all the pathogenic *E. coli* strains. ABCA12 has earlier been reported to be associated with lipid transport mechanisms, especially for the transport of glucosylceramides (Wallrodt et al., 2013). Notably, early stages of enteric infection, caused by gram-negative bacteria like *Shigella*, have been shown to be dependent on accumulation of cholesterol and sphingolipids, like glucosylceramides (Lafont et al., 2002). Given that the pathogenic factors (and the virulent mechanism) of pathogenic *E. coli* bear high resemblance to that of *Shigella*, the set of HPIs involving human ABCA12 assumes significance. Similar to ABCA12, the human protein ABCB1 was also found to interact with ChuT as well as YhdW in all the pathogenic *E. coli* strains. YhdW is a component of ABC transporter and biosynthesis of LPS in *Escherichia* and is known to be mediated by an ATP-binding cassette (ABC)-transporter-dependent mechanism (Greenfield and Whitfield, 2012). Further, an earlier study had reported the involvement of lipopolysaccharide (LPS) in adhesion, retention and transport of *E. coli* JM109 (Abu-Lail and Camesano, 2003). Moreover, a recent study has reported the role of human ABCB1 in minimizing the bacterial adhesion to the gastrointestinal cells (Crowe and Bebawy, 2012). Thus, the

predicted HPI between human ABCB1 and bacterial proteins may be crucial for deciding the fate of bacterial attachment to the host cell surface. Literature evidence supporting this small subset of HPis (which are unique to the interactome data of the studied pathogens) suggest that these HPis may be regarded as specialized adaptations of the pathogenic *E. coli* strains. These HPis probably conform a competitive edge to the pathogenic strains of *E. coli* (over the non-pathogenic strains) and may thus be implicated for their virulent nature.

Human–*E. coli* Protein Interactions Specific to Non-pathogenic Strain

The interactome data pertaining to the HPis of the non-pathogenic *E. coli* strain (MG1655) was seen to harbor a subset of 45 HPis which were found to be absent for the pathogenic strains (Figure 1). Bacterial proteins like Fe(3+) dicitrate transport protein (FecA), Fe(3+) dicitrate-binding periplasmic protein (FecB), primary amine oxidase (TynA), and D-allose-binding periplasmic protein (AlsB) were predicted to be involved in interactions with several human proteins including Ribulose-phosphate 3-epimerase (RPE), Antigen peptide transporter 2(TAP2), and a host of aldehyde dehydrogenases and ABC transporters. In complementation to the *fhu* operon, MG1655, like most other commensal and laboratory strains of *E. coli*, was observed to possess the *fec* operon, an additional mechanisms for iron uptake. FecA and FecB (members of the *fecABCDE* ferric citrate transport system) in MG1655 was seen to interact with several host (human) cell transporters including, ABCA10, ABCG8, ABCB4, ABCC6, and ABCC8. In yet another PPI that is specific to the non-pathogenic strain, AlsB, the periplasmic substrate-binding component of the allose ABC transporter, was found to interact with human RPE. It was interesting to note that the *alsBAC* system, though primarily meant for allose transport, could also function as a low affinity ribose transporter (Kim et al., 1997). The above observations lead us to hypothesize that the HPI involving AlsB and RPE in *E. coli* MG1655 is probably involved in uptake of additional nutrients to sustain its growth in a nutrient-depleted environment.

TynA, the copper-containing amine oxidase, was observed to interact with host alcohol dehydrogenases. Although, previous studies involving *E. coli* amine oxidase could not establish the role of TynA in pathogenesis, TynA activity (coupled with H₂O₂ production) has been demonstrated for growth under stringent environmental conditions (Murooka et al., 1978; Elovaara et al., 2015). Furthermore, the potential of TynA to utilize certain (unknown) substrate on human leukocytes have also been solicited (Elovaara et al., 2015). The observed PPIs between host alcohol dehydrogenases and bacterial TynA assume importance and require a deeper probe in this regard.

Overall, the non-pathogenic strain of *E. coli* was seen to be involved in a higher number of HPis as compared to the pathogenic strains. Although the exact reason behind the existence of this added repertoire of HPis in non-pathogenic *E. coli* remains unclear, it is likely that this subset of HPis facilitates the non-pathogenic strain of *E. coli* (MG1655) in optimally utilizing the resources within the nutrient-depleted environment of the host gut.

Lineage and Strain Specific Human–*E. coli* Protein Interactions

A subset of inter-species PPIs was predicted to be specifically present in the interactome data pertaining to certain lineages of *E. coli*. In one such case, YbeQ was observed to interact with several human proteins including derlin-1 (DERL1), protein OS-9 (OS9), E3 ubiquitin-protein ligase synoviolin (SYVN1), and ubiquitin-conjugating enzyme E2 J1 (UBE2J1). Interestingly, these host proteins are known to be functionally linked to the endoplasmic reticulum-associated protein degradation (ERAD) pathway. ERAD ensures that only the correctly folded proteins are secreted by the host (human) cells. Some of these secreted proteins (whose correct folding conformations are ensured by ERAD) may be detrimental to the invading pathogen. Inspection into the *E. coli* YbeQ protein sequence revealed the presence of repetitive SEL-1 domain. Evidence from literature suggests that (eukaryotic) proteins containing SEL-1 domain(s) often influences the mechanism of ERAD (Neuber et al., 2005). Consequently, the interaction of bacterial YbeQ with the host protein involved in ERAD seems to be significant and requires further investigation. This subset of interactions was observed in the host–pathogen interactome data corresponding to only two (EDL933 and Sakai) out of the four *E. coli* O157:H7 strains as well as in *E. coli* K-12 strain. Interestingly, EDL933 and Sakai are members of lineage I of *E. coli* O157:H7 which are known to be better equipped in colonizing the human gut as compared to the other lineages of *E. coli* O157:H7 (Lowe et al., 2009). Similar HPis, specific to members of lineage I of *E. coli* O157:H7, involving a poorly characterized bacterial ferric transporter protein (herein referred to as ucp1) and host proteins like ABCA12, ABCB1, ABCB4, and ABCB5 were also observed. Such HPis suggest possible roles of YbeQ and ucp1 in manipulating the host immunity and/or metabolism, thereby aiding in preferential colonization of the human gut. Furthermore, a HPI involving bacterial FdoG (α subunit of formate dehydrogenase O) and human EEF1A2 (Elongation factor 1-alpha 2) was observed to be unique to *E. coli* TW14359 strain. Interestingly, FdoG mutant *Escherichia* and *Salmonella* have previously been shown to be susceptible to human bactericidal/permeability-increasing proteins (BPIs) during stationary phases of growth (Barker et al., 2000). This underlines the key role of bacterial FdoG in evading host defenses, thereby leading to sustained enteric infection. However, the specific mechanism as well as consequence of the interaction between EEF1A2 and FdoG is not clear.

In spite of its non-pathogenic nature, *E. coli* MG1655 was seen to share at least six PPIs with *E. coli* EDL933, a pathogenic strain of lineage I, which was found to be absent in one or more of the other studied pathogenic strains. Of them, the interaction between bacterial YcjM and human PYGB was found to be absent in all except *E. coli* MG1655 and EDL933 strains. Interestingly, PYGB has been seen to be up-regulated in studies pertaining to pulmonary infections caused by *Burkholderia cepacia* (Mariappan et al., 2013). The study however have not introspected the probable reason behind such observation. Given the key role of PYGB in glycogen metabolism, the observed HPI may have significant bearing on the host metabolism. Thus, our

finding opens up an opportunity to further investigate into the mechanism of HPIs involving human PYGB.

Interaction among *E. coli* UidC and human GUSB is yet another example of a HPI that was observed to be selectively present in host-pathogen interactome data pertaining to *E. coli* MG1655, EDL933, and TW14359. The UidC in *E. coli* is a component of the *uidABC* system, which is involved in transport and utilization of glucuronides. It is pertinent to note that steroid hormones, pharmaceuticals, toxins, xenobiotics, etc. are often excreted by the host (human) liver through bile as conjugated glucuronides. These glucuronides may be deconjugated by bacterial glucuronidases (like UidA), to release the apolar aglycone which in turn, may be utilized by the bacteria as carbon source (Liang et al., 2005). Notably though, the *uidABC* system is known to be non-functional in MG1655 due to a mutation in the *uidB* gene (Liang et al., 2005). These subtle differences which were observed in the HPI profiles among the different strains of *E. coli* may probably be attributed to evolutionary events like speciation, horizontal gene transfer, gene loss, as well as periodic insertion-deletion-substitution of nucleotides, thereby leading to lineage and/or strain specific variations.

Understanding the Role of Human and *E. coli* Proteins during Infection

The analysis of the intra-species PPIs involving different strains of *E. coli* revealed that the majority of the PPIs are common to both the pathogenic as well as the non-pathogenic strains. Only a small subset of PPIs was predicted to differentiate the pathogenic strains from the non-pathogenic ones. It would therefore be interesting to understand whether this small subset of interaction plays a role in the manifestation of *E. coli* infection process. The KEGG pathway for infections caused by pathogenic *E. coli* (ko05130), enlists the human proteins which are probably involved in the virulence process. We investigated the human proteins, which are reported (in the KEGG database) to be involved in the *E. coli* infection machinery (referred to as set-B in the Materials and Methods Section) to check whether they are connected to the host (human) protein associated to PPIs with *E. coli* (referred to as set-A in the Materials and Methods Section). We observed that majority of the proteins of human origin, which were reported (in KEGG) to be affected by the *E. coli* pathogenesis, were closely related to the human proteins participating in PPIs (Figure 2). Surprisingly though, for all the five studied *E. coli* strains, the list of human proteins which were predicted to take part in PPIs, and were also closely associated with the proteins involved in *E. coli* infection machinery (as reported by KEGG) was identical. This probably indicate that the small subset of PPIs which were specific to either of the pathogenic or the non-pathogenic strains do not play a direct role in the *E. coli* infection machinery. However, given that most studies related to *E. coli* are based on the laboratory cultured strains (like *E. coli* MG1655), it is probable that our observation is an effect of the lack of proper understanding of the *E. coli* infection machinery. While human proteins (reported in ko05130) like CDC42, CD14, PRKCA, TLR4, etc. were directly

connected (first degree neighbors) to one or more of the host proteins participating in PPIs, others proteins reported in the KEGG pathway for *E. coli* virulence (such as, NCK2, LY96, ABL1, and CDH1) were observed to be close neighbors (second degree connections) of the interacting host proteins (data not shown). Some of the human proteins (viz. NOS3, MAPK14, TNF, NFKB1) which were predicted to be involved in PPIs, were observed to be directly linked (first degree connection) to multiple host proteins from the KEGG *E. coli* virulence pathway (ko05130) including RHOA, CD14, PRKCA, TLR4, and CDC42. Notably, the mechanism by which the *E. coli* proteins influence the functionality of the human proteins (as reported in ko05130) is not clearly known. For example, although it is known that the *E. coli* proteins are involved in altering the host processes like cytoskeletal rearrangement, microtubule destabilization and programmed cell death (apoptosis), the mechanistic aspect of how the components of the type III secretion system transmute these biological processes of the host is poorly characterized. The above observations motivated us to inspect into the possible molecular mechanisms that may be involved in the cellular signaling processes that are triggered in response to the pathogenic *E. coli* proteins.

It is evident from KEGG's Pathogenic *E. coli* Infection pathway (ko05130), that there is an inter-play between the flagellar assembly/biogenesis process of EHEC and the human Toll-like receptor signaling pathway, which finally leads to host inflammatory responses. Signaling through the Toll-like receptor signaling pathway is triggered by TLR4 and TLR5 (Figure 2). However, the mechanism of action is not clearly understood. Our analysis revealed that the human TLR4 is functionally linked to MAPK14, which in turn, interact with the pathogenic DegS (Supplementary Table 1). The bacterial DegS is a member of the DegU-DegS two component system which is known to be associated with flagellary movement. A recent study on *Bacillus subtilis*, has reported that the inhibition of flagellar rotation functions acts as a mechanical trigger for the activation of the DegU-DegS two-component signal transduction system (Cairns et al., 2013). Once the pathogenic *E. coli* attaches to any surface (such as, the human intestinal epithelial cell surface), the flagellary motor movement is halted, thus triggering DegU-DegS system and thereby allowing DegS to interact with the human proteins. This indicates that the mechanism of activation of the host Toll-like receptor signaling pathway is possibly mediated through the interactions between the *E. coli* DegS and the human MAPK14 and TLR4 proteins. It is also interesting to note that DegU-DegS is also responsible for the biosynthesis of the exopolymer poly- γ -dl-glutamic acid, which is associated with immune evasion and virulence in *Staphylococcus epidermidis* (Kocianova et al., 2005). Furthermore, TLR4 and TLR5 are also observed to interact with CDC42, a protein involved in regulation of actin cytoskeleton assembly. Given this role of CDC42 in microtubule rearrangement, a probable role of DegS in mediating phagocytosis may also be speculated.

The most common symptom of diseases caused by the enteropathogenic *E. coli* is watery diarrhea. Disruption of tight junctions and monolayer permeability leading to loss of fluids

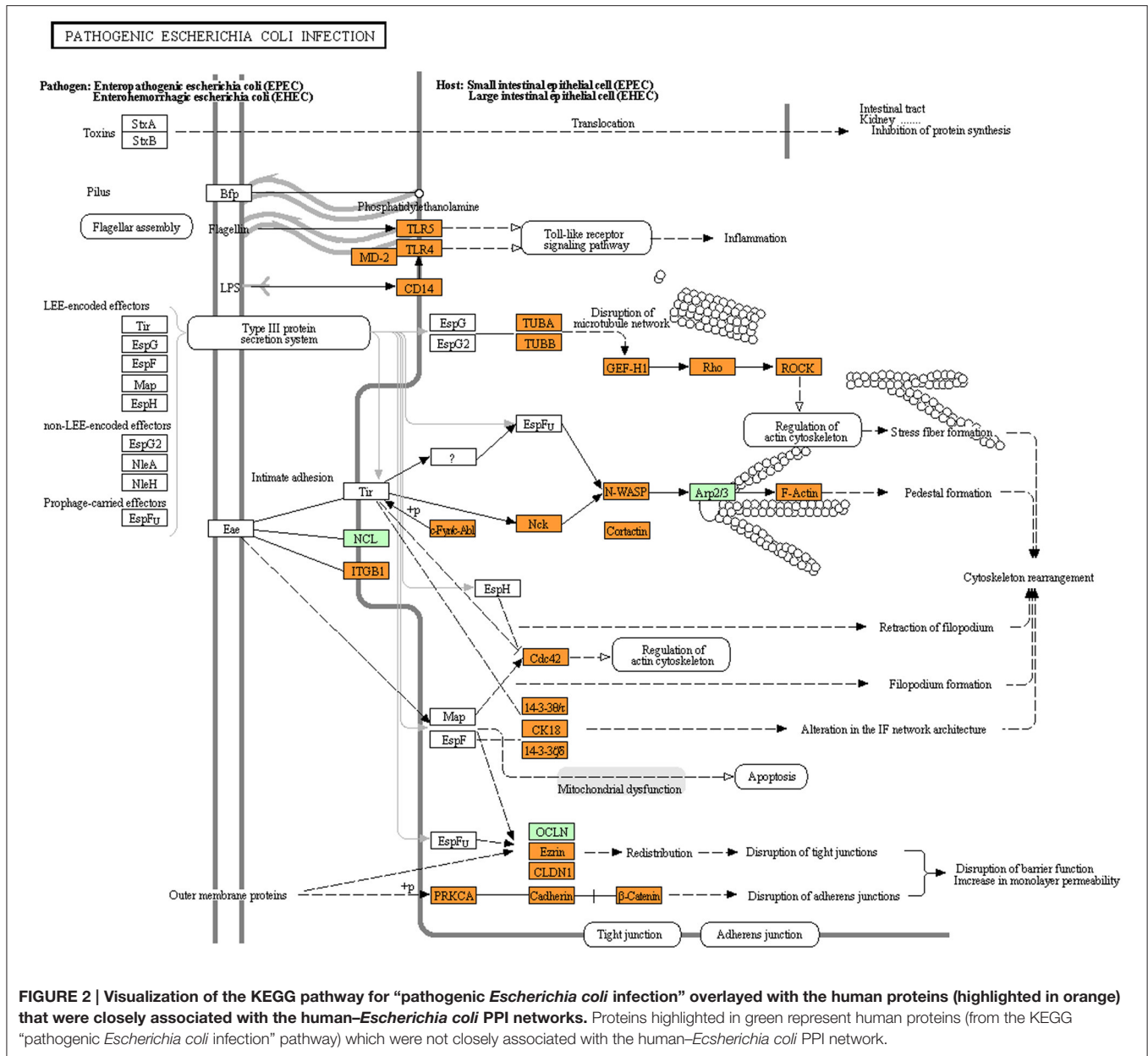


FIGURE 2 | Visualization of the KEGG pathway for “pathogenic *Escherichia coli* infection” overlaid with the human proteins (highlighted in orange) that were closely associated with the human-*Escherichia coli* PPI networks. Proteins highlighted in green represent human proteins (from the KEGG “pathogenic *Escherichia coli* infection” pathway) which were not closely associated with the human-*Escherichia coli* PPI network.

from epithelial cells has been reported to be the cause of diarrhea. E-cadherin and its interaction with beta-catenin (CTNNB1) are known to regulate the monolayer permeability of the host cells. It has been reported that once the pathogenic *E. coli* adhere to the host epithelial cells, its outer membrane proteins (OMPs) dissociate the cadherin/beta-catenin complex, thereby leading to watery diarrhea (Malladi et al., 2004). Adhesion of *E. coli* to the epithelial cell membrane is however a pre-requisite for the cause of any such clinical symptoms. Mannose and its derivatives like mannose-6-phosphate are known to inhibit the process of cellular adhesion by the pilated strains of *E. coli* to the mammalian epithelial cells (Ruggieri et al., 1985). In this regard, it is pertinent to note the involvement of pathogenic phosphomannose isomerase (PMI/ManA) in HPI with the human proteins. ManA, which is secreted by *E. coli*, catalyzes the

conversion of mannose-6-phosphate and fructose-6-phosphate, thereby, allowing the *E. coli* to adhere to the host epithelial cells (Ruggieri et al., 1985). In addition, in studies relating to human breast cancer cell lines, phosphoglucose isomerase (PGI), which is a functional analog of ManA/PMI, has been seen to down regulate E-cadherin/beta-catenin complex formation through inhibition of E-cadherin expression and enhancement of proteosomal degradation of beta-catenin (Wu et al., 2008; Funasaka et al., 2009; Jia et al., 2013). These observations indicate that the bacterial ManA/PMI may play a very crucial role in pathogenesis. The above evidence seems to suggest that in addition to aiding the bacteria to adhere to the host cell surface, ManA may also aid the pathogen in disrupting the tight junctions within the host epithelial cell lining, thereby resulting in watery diarrhea.

Potential of *E. coli* to Colonize within the Human Gut

HPIs play a critical role in the survival and sustenance of external agents like viruses and bacteria inside a host system. The HPIs help the invading pathogen to adhere to the host cell surface and subsequently to counter the host's immune responses. In addition, in cases of intracellular pathogens like *Mycobacterium tuberculosis*, the HPIs also play part in facilitating the phagocytic uptake (by host immune cells) and intra-cellular survival of the pathogen. Notably, it was observed that the majority of inter-species PPIs were ubiquitous to all the studied strains of *E. coli*, irrespective of their pathogenic nature. Further, the human proteins involved in PPIs and those which are closely associated with the *E. coli* infection machinery (as reported in KEGG database) were mostly similar for all the studied strains of *E. coli*. Given this, the gene expression profiles of the human and *E. coli* proteins which were involved in HPIs were studied to estimate the proportion of proteins which may be involved in these interactions. It was assumed that a ratio of cumulative score of bound protein to total protein concentrations (involved in different PPIs) would indicate the extent of colonization. Based on this assumption, a metric to determine the host colonization potential of the *E. coli* strains was designed. This metric was termed as the HCI. The procedure of deriving the HCI metric has been described in the Materials and Methods Section.

It was observed that host-associated stains of *E. coli* had higher HCI-values (>0.99) as compared to the non-pathogenic laboratory strain (0.823), which is indicative of the superior potency of host-associated (pathogenic) strains to colonize the human gut. Furthermore, it was observed that strains of both lineages I and lineages I/II had comparable HCI-values (0.996 and 0.997, respectively). It may be noted that strains of lineages I (EDL933 and Sakai) and lineages I/II (EC4115 and TW14359) had fewer number of PPIs as compared to laboratory (or environmental) strains. This indicated that HCI is not dependent on the total number of PPIs but is a function of the amount of proteins that are produced in the cell and are involved in the PPI events. Thus, host–pathogen pair which share a fewer number of interactions, can potentially have a relatively higher HCI as compared to a host–pathogen pair sharing a larger array of PPIs, and subsequently a better colonization potential. Pertinently, unlike gram-negative bacteria such as *E. coli*, the gram-positive bacteria possess a thicker cell wall and harbors a significantly lower number of outer membrane vesicles. Hence, in the case of gram-positive bacteria, a fewer number of surface (and secreted) proteins are expected to take part in interaction with the host. HCI, being independent of the total number of PPIs, is aptly suited to quantitatively portray the potency of a pathogen to successfully survive within the host.

Comparison between Host–*E. coli* Interaction Patterns in Humans and Cattle

E. coli O157:H7 is primarily known to colonize the ruminant gastrointestinal tract (Zhang et al., 2010). Interestingly though, while human infected with *E. coli* O157:H7 suffers from several

enteric diseases, cattle infected with *E. coli* O157:H7 appear to be mostly asymptomatic (Karmali, 1989; Karch et al., 2005; Zhang et al., 2007). As a result, some of the earlier studies were focused on identifying the spatiotemporal nature of immune response in cattle when infected with *E. coli* O157:H7 (Tildesley et al., 2012). In an attempt to get better insights into the basis of the variations in immune responses of humans and cattle on exposure to *E. coli* O157:H7 infection, analysis of the HPI profiles of both humans and cattle was carried out. For this purpose, the profiles pertaining to HPIs between cattle and the five studied strains of *E. coli* were obtained using a strategy similar to that adopted while predicting HPIs between humans and *E. coli*. Homologs between the human and cattle proteins were determined using a BLASTp analysis ($e\text{-value} \leq 1 \times e^{-10}$, identity $\geq 60\%$ and minimum overlap of 80%). **Supplementary Figure 1** presents a comparative analysis of HPI profiles of humans and cattle pertaining to the five studied strains of *E. coli*. Overall, humans seemed to be involved in a higher number of HPIs with *E. coli* than cattle. Further, in spite of certain commonalities among the HPI profiles of the two hosts, the majority of the interactions of human and cattle (with each of the studied strains of *E. coli*) were largely different. Given that in most cases cattle infected with *E. coli* O157:H7 are phenotypically asymptomatic, the above observation is important. A deeper probe into the functional role of the *E. coli* proteins involved in these differential HPIs among humans and cattle could help in understanding the basis of infections caused in humans.

In order to ascertain the exact role of the *E. coli* proteins which were observed to participate in HPIs with the host, a Gene Ontology (GO) Enrichment Analysis of these proteins (and their first degree neighbors) was performed. Overall, biological processes associated with transport of nutrients like ion transport (GO:0006811), amino acid transport (GO:0006865), carboxylic acid transport (GO:0046942), organic acid transport (GO:0015849), siderophore transport (GO:0015891), etc. were found to be enriched in all the sets of *E. coli* proteins involved in HPIs (irrespective of the associated host), suggesting the need for *E. coli* to acquire nutrients from its host for its survival and proliferation. GO terms were further inspected to identify biological process which was unique to the *E. coli* protein involved in HPI with humans and cattle (**Table 1**). Notably, the *E. coli* proteins associated with HPI in human were uniquely enriched in carbohydrate metabolic process terms. These included monosaccharide metabolic process (GO:0005996), cellular carbohydrate metabolic process (GO:0044262), glycolysis (GO:0006096), hexose metabolic process (GO:0019318), monosaccharide catabolic process (GO:0046365), and cellular carbohydrate catabolic process (GO:0044275). In contrast, *E. coli* proteins associated with HPI in cattle were uniquely enriched in nucleotide and nitrogen metabolic process terms, such as, nucleotide biosynthetic process (GO:0009165), nucleobase metabolic process (GO:0009112), purine salvage (GO:0043101), nitrogen compound biosynthetic process (GO:0044271). It is pertinent to note that earlier literatures have pointed at the vital role of gut micro-flora in production of amino acids (and subsequent protein assimilation) inside the rumen gut (Genzebu and Tesfay, 2015).

TABLE 1 | Uniquely enriched Gene Ontology (GO) terms corresponding to the proteins from the five studied strains of *Escherichia coli* which participated in PPI with humans and cattle.

| GO terms associated with <i>E. coli</i> proteins specifically interacting with human | GO terms associated with <i>E. coli</i> proteins specifically interacting with cattle |
|--|---|
| GO:0005996—monosaccharide metabolic process | GO:0055086—nucleobase, nucleoside, and nucleotide metabolic process |
| GO:0001539—ciliary or flagellar motility | GO:0034654—nucleobase, nucleoside, nucleotide, and nucleic acid biosynthetic process |
| GO:0048870—cell motility | GO:0009165—nucleotide biosynthetic process |
| GO:0044262—cellular carbohydrate metabolic process | GO:0009112—nucleobase metabolic process |
| GO:0006096—glycolysis | GO:0006163—purine nucleotide metabolic process |
| GO:0019318—hexose metabolic process | GO:0044271—nitrogen compound biosynthetic process |
| GO:0032774—RNA biosynthetic process | GO:0043101—purine salvage |
| GO:0046365—monosaccharide catabolic process | GO:0043174—nucleoside salvage |
| GO:0044275—cellular carbohydrate catabolic process | GO:0042946—glucoside transport |
| | GO:0015772—oligosaccharide transport |

More interestingly, the human interactome components of *E. coli*, were also seen to be enriched in biological processes pertaining to ciliary or flagellar motility (GO:0001539), and cell motility (GO:0048870). Proteins associated with these processes were found to be absent in the HPI profile pertaining to cattle. Given that ciliary or flagellar motility functions may grossly be associated to virulence, this observation is particularly significant, especially in perspective of the fact that cattle harboring pathogenic strains of *E. coli* (like the O157:H7 strain) in their gut are mostly asymptomatic in nature. The human protein which were observed to interact with the *E. coli* proteins associated to ciliary and flagellar movement were seen to be involved in lipid transport (GO:0006869), sterol transport (GO:0015918) and transmembrane transport (GO:0055085) functions (data not shown). This is in concordance with the recent reports which highlighted the role of lipids residing on the host cell surface in flagella mediated bacterial adhesion to host epithelium (Rossez et al., 2014, 2015).

DISCUSSION

The current study encompasses the intra-species PPI data from human as well as different strains of *E. coli* to infer potential HPis among the host and the pathogen proteins. Results indicate that the invading pathogens have devised ways to subvert the host's immune defenses, primarily by countering the redox stresses generated by the host. All the studied strains of *E. coli* were observed to possess an arsenal of genes including CysP, GadA, KatG, and KatE to

neutralize the oxidative stress that are generated by the host's immune system in response to the pathogenic determinants in *E. coli*.

Different strains of *E. coli* were also observed to harness specialized repertoire of genes to acquire “scarce” nutrients from its environment. For example, FhuA and FhuD (members of the *fhu* operon) which were observed to interact with host ABC transporters are essential for iron acquisition for the pathogen under *in vivo* conditions. Certain non-pathogenic strains of *E. coli* (like *E. coli* K-12 MG1655) encoded for additional mechanisms of iron uptake through the ferric citrate transport system (*fec* operon). Furthermore, it was also interesting to note the selective presence of *uidABC* system in a few strains of *E. coli*. The *uidABC* system is involved in the deconjugation of glucuronides and subsequent uptake of the apolar aglycone by the bacteria to be used as carbon source for its growth. The *uidABC* system of MG1655 is particularly fascinating in this regard. Although the genome of MG1655 encodes for all the genes constituting the *uidABC* system, the operon is non-functional due to a single mutation at position 100 of the *uidB* gene (Liang et al., 2005), implying that the *E. coli* K-12 MG1655 is unable to utilize the rich bounty of glucuronides that are available inside the host gut. This leads us to speculate that one or more of the glucuronides (in its aglycone form), which often include steroid hormones, pharmaceuticals, toxins, xenobiotics, etc., may be toxic for the bacteria and most strains of *E. coli*, including the K-12 strains lacking the capacity to detoxify the same.

Genomes of the pathogenic strains of *E. coli* were seen to encode for additional proteins which were involved in the infection process. For example, pathogenic strains of *E. coli* encoded for HylA and EivG which aided the pathogen in biofilm formation and secretion respectively. In addition, the pathogenic strains also possessed genes like YhdW which helped the pathogen in transporting lipopolysaccharide (LPS) to the cell surface. Notably, LPS is known to help the bacteria in adhering to the host epithelium (Abu-Lail and Camesano, 2003). Consequently, given the role of human ABCB1 in minimizing bacterial adhesion to the gastrointestinal epithelium (Crowe and Bebawy, 2012), the predicted PPIs involving YhdW and human ABCB1 assumes importance.

Interestingly, there were no differences among the set of human proteins which interacted with different strains of *E. coli* and were in close proximity to those proteins which are known to be affected by *E. coli* pathogenesis (from the KEGG database). It may be speculated that there are additional factors like altered gene expression, alternate gene splicing, and epigenetic modifications which play a decisive role during the pathogenesis process.

Further, a comparative analysis of the HPis pertaining to different strains of *E. coli* showed lineage and clade specific differences. Analysis of the enriched GO terms corresponding to the *E. coli* proteins (which were involved in HPis), revealed that members of *E. coli* O157:H7 lineage I were uniquely enriched in lipopolysaccharide biosynthetic process (GO:0009103), response to hydrogen peroxide (GO:0042542), response to reactive oxygen species (GO:0000302), etc. This is particularly significant in

the light of the fact that previous literature suggests that the *E. coli* O157:H7 strains belonging to lineage I are better adapted to survive within the human gut as compared to the other strains of *E. coli* O157:H7 (Lowe et al., 2009). Furthermore, genes/proteins involved in the GABA dependent acid-resistance were observed to be specifically enriched in *E. coli* Sakai strain. The above findings indicate that different strains of *E. coli* O157:H7 have evolved (or acquired) certain strain/clade/lineage specific techniques to bypass the host defenses.

In spite of the minor variations in the HPI profiles, the different biological pathways that are associated with the human proteins interacting with the different strains of pathogenic and non-pathogenic *E. coli* were observed to be largely similar. Yet, we know that the virulent potential of these pathogenic strains are not the same. These lead us to speculate that there might be strain specific variations in the level of expression of some of the pathogenic genes/proteins. In this regard, it was interesting to note the levels of gene expression of the stress responsive genes among *E. coli* O157:H7 strains belonging to different lineages (Wang et al., 2009). For example, on exposure to peroxynitrate stress, while the *cysP* gene, was observed to be significantly over-expressed among members of lineage I/II (TW14359) as compared to those of lineage I (Sakai). However, no significant variations in gene expression profile were observed for the other stress responsive genes (which participated in PPI with the host proteins; Wang et al., 2009). This indicates that in spite of possessing a common arsenal of stress responsive genes, *E. coli* residing in different ecological niches is able to alter the expression of these genes to efficiently cope with the environmental stresses.

Given the importance of the PPIs in the colonization/survival of the bacteria inside the host, it was assumed that the amount of host and bacterial proteins which was involved in these interactions would be indicative of the colonization potential of the bacteria. Thus HCI, the novel metric for estimating host colonization potential has been proposed. Host-associated (pathogenic) strains of *E. coli* were seen to demonstrate higher HCI-values as compared to the laboratory (non-pathogenic) strain. Higher HCI-values of the O157:H7 strains are indicative of their superior potential to colonize the human gut. In comparison, strains belonging to different lineages of *E. coli* O157:H7 had similar HCI-values, indicating that in spite of the differences in their virulent potential, their ability to colonize the human gut is the same.

The identified HPI are expected to be helpful in deciphering the biological basis of the diseases caused by different bacterial pathogens including *E. coli*. Investigating time-series gene expression data for the host and pathogen protein involved in HPIs would be beneficial in understanding the chronology of events during infection. The study opens up the scope for identifying new therapeutic targets in enterohemorrhagic pathogens. It also provides the potential of developing host directed therapies which are highly solicited in this era of rapid emergence of drug-resistant traits among different strains of bacteria.

MATERIALS AND METHODS

Data Acquisition and Preprocessing

A pre-requisite for comparative analysis of HPIs pertaining to different strains of *E. coli*, was to identify orthologous groups of protein among the studied strains. In addition, computation of host-pathogen interacting pairs also necessitated the determination of homologous protein sequences in humans and *E. coli*. Consequently, the protein sequence information for the five studied strains of *E. coli* as well as that of human were downloaded from the NCBI database (<ftp://ftp.ncbi.nlm.nih.gov/genomes/>). The protein sequences from the five strains of *E. coli* were subsequently clustered by employing the BLASTClust program from the standalone BLAST package. A total of 6,884 unique clusters, each representing a distinct functional group, were obtained using threshold criteria of 95% identity and 80% overlap among the clustered sequences. The sequence identity and overlap criteria were obtained from previous literatures on protein clustering (Junjie et al., 2010; Finn et al., 2014). The homologous sequences between *E. coli* and humans were identified by BLASTp analysis. Human and *E. coli* proteins were considered to be homologous if they shared at-least 30% identity and 80% sequence overlap with an e-value lower than $1 \times e^{-10}$.

The intra-species interactome data for human and the five studied strains of *E. coli*, which were considered for creating the background library of template interactions, were downloaded from the STRING database, version 9.1 (<http://string-db.org/>; Franceschini et al., 2013). It is important to note that, in addition to the experimentally validated PPIs, the STRING database also collates PPI information using a wide range of techniques including text mining, homology determination, etc. PPIs determined by such non-validated techniques may often be spurious and are typically characterized by low combined confidence score. To minimize the false discovery rate of predicted HPIs, only the intra-species PPIs with a combined confidence score of 0.9 or higher were retained in the final set of template interactions. It is also pertinent to note that the STRING data classifies any given strain of bacteria as either “core” or “periphery.” It was observed that the PPI interactome data for the “periphery” strains were incomplete. Further, all the studied *E. coli* strains with the exception of *E. coli* MG1655 were found to be classified as “periphery” strains in the STRING database. Consequently, an interlog based mapping approach was put to use in order to obtain the complete interactome pertaining to all the five strains of *E. coli*. The interlog approach (Yu et al., 2004) is explained in the context of the current study in **Figure 3A**. The intra-species PPIs that were thus identified by this interlog mapping approach, were combined with the PPI data obtained from the STRING database (filtered with a combined confidence score of 0.9 or higher) to obtain the final set of template interactions.

Identification of Sub-cellular Localization of Proteins

Determining the sub-cellular localization of proteins is an important step in identifying HPIs. Sub-cellular localization of a

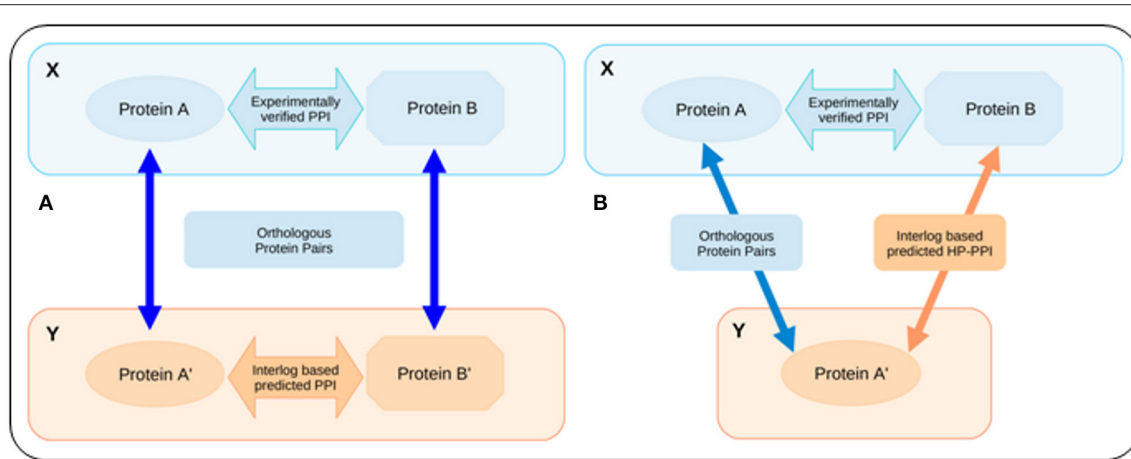


FIGURE 3 | Schematic representation of the interlogs approach. (A) Interlog based mapping approach for predicting intra-species PPIs. If A–A' and B–B' are orthologous sets of protein in two organisms X and Y and the proteins A and B are known to interact in organism X, then the interlog based mapping approach concludes a possible PPI between the proteins A' and B' in organism Y. **(B)** Modified (inter-species) interlog method that was employed for predicting the HP-PPIs. X represents the host and Y represents the pathogen (or vice versa). If protein A and protein B of organism X are known to interact among each other, and if there exists a homolog to protein A (protein A') in organism Y, then the modified interlog method assumes that protein A' will also interact with protein B through a PPI.

protein governs its feasibility of participate in a HPI. For example, a pathogenic protein localized on the outer membrane or cell wall or one which gets secreted has a better chance of interacting with host proteins as compared to a pathogenic protein that is localized within the cytoplasm. Similarly, a membrane-associated or secreted host protein is more likely to interact with the proteins of an extra-cellular pathogen (like most species of *E. coli*). Since we were unable to mine the sub-cellular localization of most of the studied proteins from literature, we resorted to *in silico* prediction techniques. The sub-cellular localization of the bacterial proteins was ascertained using PSORTb (ver.3.0.2). PSORTb employs a signal sorting based search technique and was observed to outperform most of its competitors (Yu et al., 2010). Sub-cellular localization for the human proteins was obtained using WoLF PSORT, a widely used tool for eukaryotic protein sub-cellular localization prediction (Horton et al., 2007). The predicted sub-cellular localization of proteins from both the humans (host) and the five strains of *E. coli* (pathogen) were utilized in later steps for filtering the predicted (potential) HP-PPIs.

Prediction of Host–*E. coli* Protein–Protein Interactions

The PPIs were predicted using an approach similar to the interlogs method (Yu et al., 2004). It stated that if a PPI is known to exist among any two proteins in an organism and there are orthologs of those two proteins in any other organism, then those orthologous proteins should also participate in a PPI. HP-PPIs (inter-species interlogs) were identified on the basis of the (intra-species) template interactions and the sequence homology data (that was obtained using the BLASTp analysis between human and *E. coli* proteins). In the inter-species interlogs method, a HPI was assumed to exist if (a) for a pathogen-specific (intra-species) template PPI, there is a homolog of at-least one of the interacting proteins in the host; or (b) for a host-specific

(intra-species) template PPI, there is a homolog of at-least one of the host proteins in the pathogen. **Figure 3B** depicts a schematic diagram of the modified (inter-species) interlogs approach that was employed for determining the HP-PPIs (inter-species interlogs). The set of potential HP-PPIs (corresponding to the five studied strains of *E. coli*) thus identified were subsequently screened for the sub-cellular localization of the participating host and pathogen proteins. Only the subset of PPIs where both the host and the pathogen proteins were localized either on the cell membrane/cell wall or exterior to it, were considered to be legitimate host-pathogen interacting pairs. The above step resulted in the generation of five sets of host-pathogen interactome data, each corresponding to the HP-PPIs predicted among humans and one of the five studied strains of *E. coli*.

Host–*E. coli* Interaction Network Analysis

The host–pathogen interactome data pertaining to the five studied strains of *E. coli* were analyzed for various network properties like degree and betweenness centralities of each node (constituent proteins in the network). For the purpose, the host–pathogen interactome data for each of the five strains of *E. coli* was analyzed using Cytoscape (version 2.8), a widely used platform for network analysis and visualization (Shannon et al., 2003). In addition, a comparative study of the five HPI networks was also performed. CompNet (Kuntal et al., 2016), a tool for comparative analysis of multiple biological networks, was employed for the purpose.

Functional Pathway Analysis

The pathogenic *E. coli* infection pathway (ko05130) in the KEGG database enlists the human proteins which may be affected by the virulence mediated by pathogenic strains of *E. coli*. In order to determine the functional relationship among the set

of human proteins that were predicted to interact with the pathogenic proteins (set-A) and the set of human proteins which were reported in the KEGG database to be affected by *E. coli* pathogenesis (set-B), the KEGG pathway analysis was performed. The human interactome data (as obtained from the STRING database) was filtered for a combined confidence score of 0.9 or higher. Based on the filtered human PPI data, the shortest paths connecting each of the components of set-A to those of set-B were determined. Shortest paths connecting the components of set-A and set-B with a path-length of two or less were retained. The shortest paths, thus determined, helped us in identifying the first and second degree connections between the set of human proteins that were predicted to interact with the pathogenic proteins and the set of human proteins which were reported in the KEGG database to be affected by *E. coli* pathogenesis.

Computing Host Colonization Index (HCI)

Host Colonization Index (HCI) was computed as the ratio of the cumulative function of bound:total protein concentrations (involved in different HPIs). First, the gene expression values for the host and *E. coli* proteins were obtained from NCBI GEO database¹ (Gene Expression Omnibus; Maurer et al., 2005; Ukena et al., 2005; Lee et al., 2014). It was assumed that the gene expression values would be roughly proportional to the protein expression values and were considered as total concentrations (P_t) of the proteins. Next, for each of the proteins (P) which participated in HPI, the PPI module of the Steady State Gene Expression Simulator (SSGES; Srinivasan and Venkatesh, 2014) was employed to estimate the amount of host and pathogen proteins that may be bound to (interacting with) other proteins (P_b). Finally, HCI was computed as a ratio of the cumulative functions of bound:total protein concentrations for each of the host and pathogen proteins involved in HPIs.

$$HCI = \sum \frac{P_b}{P_t}$$

¹NCBI GEO database. Available online at: <http://www.ncbi.nlm.nih.gov/geo/> (Accessed March 16, 2016).

REFERENCES

- Abu-Lail, N. I., and Camesano, T. A. (2003). Role of lipopolysaccharides in the adhesion, retention, and transport of *Escherichia coli* JM109. *Environ. Sci. Technol.* 37, 2173–2183. doi: 10.1021/es026159o
- Barker, H. C., Kinsella, N., Jaspe, A., Friedrich, T., and O'Connor, C. D. (2000). Formate protects stationary-phase *Escherichia coli* and *Salmonella* cells from killing by a cationic antimicrobial peptide. *Mol. Microbiol.* 35, 1518–1529. doi: 10.1046/j.1365-2958.2000.01820.x
- Branchu, P., Matrat, S., Varelle, M., Garrivier, A., Durand, A., Crépin, S., et al. (2014). NsrR, GadE, and GadX interplay in repressing expression of the *Escherichia coli* O157:H7 LEE pathogenicity island in response to nitric oxide. *PLoS Pathog.* 10:e1003874. doi: 10.1371/journal.ppat.1003874
- Cairns, L. S., Marlow, V. L., Bissett, E., Ostrowski, A., and Stanley-Wall, N. R. (2013). A mechanical signal transmitted by the flagellum controls signalling in *Bacillus subtilis*. *Mol. Microbiol.* 90, 6–21. doi: 10.1111/mmi.12342

GO Enrichment Analysis

The Gene Ontology (GO) enrichment analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) Functional Annotation Chart tool (Huang et al., 2009a,b). GO term enrichment analysis (level 4 BP terms) of the *E. coli* proteins involved in HPIs was performed with the following threshold cut-offs: count ≥ 2 ; and $p < 0.01$ (Bonferroni corrected).

AUTHOR CONTRIBUTIONS

TB, KV, and SM conceptualized the work and the design of experiments. TB optimized critical parameters for conducting the experiments with inputs from KV and SM. TB performed the experiments, analyzed the results, and wrote the paper. KV and SM edited the paper.

FUNDING

This work was partially supported by grants from the Department of Biotechnology (DBT), Government of India (BT/PR3260/BRB/10/967/2011).

ACKNOWLEDGMENTS

We express our sincere gratitude to Dr. Sumana Srinivasan, who helped us with the technicalities of using the SSGES tool for computing the HCI. We also thank Dr. Anirban Dutta and Ms. Chandrani Das for several helpful discussions and comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fcimb.2017.00128/full#supplementary-material>

Supplementary Table 1 | List of the predicted inter-species PPIs involving human proteins and those from the five studied strains of *Escherichia coli*.

Supplementary Figure 1 | Comparison analyses of the inter-species PPI patterns in humans and cattle involving proteins from different *Escherichia coli* strains.

- Corbishley, A., Ahmad, N. I., Hughes, K., Hutchings, M. R., McAteer, S. P., Connelley, T. K., et al. (2014). Strain-dependent cellular immune responses in cattle following *Escherichia coli* O157:H7 colonization. *Infect. Immun.* 82, 5117–5131. doi: 10.1128/IAI.02462-14
- Crowe, A., and Bebawy, M. (2012). ABCB1 (P-glycoprotein) reduces bacterial attachment to human gastrointestinal LS174T epithelial cells. *Eur. J. Pharmacol.* 689, 204–210. doi: 10.1016/j.ejphar.2012.05.047
- Deng, K., Wang, S., Rui, X., Zhang, W., and Tortorello, M. L. (2011). Functional analysis of ycfR and ycfQ in *Escherichia coli* O157:H7 linked to outbreaks of illness associated with fresh produce. *Appl. Environ. Microbiol.* 77, 3952–3959. doi: 10.1128/AEM.02420-10
- Elovaara, H., Huusko, T., Maksimow, M., Elima, K., Yegutkin, G. G., Skurnik, M., et al. (2015). Primary amine oxidase of *Escherichia coli* is a metabolic enzyme that can use a human leukocyte molecule as a substrate. *PLoS ONE* 10:e0142367. doi: 10.1371/journal.pone.0142367

- Eppinger, M., Mammel, M. K., Leclerc, J. E., Ravel, J., and Cebula, T. A. (2011). Genomic anatomy of *Escherichia coli* O157:H7 outbreaks. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20142–20147. doi: 10.1073/pnas.1107176108
- Evans, P., Dampier, W., Ungar, L., and Tozeren, A. (2009). Prediction of HIV-1 virus-host protein interactions using virus and host sequence motifs. *BMC Med. Genomics* 2:27. doi: 10.1186/1755-8794-2-27
- Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223
- Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., et al. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 41, D808–D815. doi: 10.1093/nar/gks1094
- Funasaka, T., Hogan, V., and Raz, A. (2009). Phosphoglucose isomerase/autocrine motility factor mediates epithelial and mesenchymal phenotype conversions in breast cancer. *Cancer Res.* 69, 5349–5356. doi: 10.1158/0008-5472.CAN-09-0488
- Genzebu, D., and Tesfay, G. (2015). The role of bacteria in nitrogen metabolism in the rumen with emphasis of cattle. *Res J Agri Env Mgmt.* 4, 282–290.
- Goldberg, R. F., Austen, W. G., Zhang, X., Munene, G., Mostafa, G., Biswas, S., et al. (2008). Intestinal alkaline phosphatase is a gut mucosal defense factor maintained by enteral nutrition. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3551–3556. doi: 10.1073/pnas.0712140105
- Greenfield, L. K., and Whitfield, C. (2012). Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways. *Carbohydr. Res.* 356, 12–24. doi: 10.1016/j.carres.2012.02.027
- Henard, C. A., and Vázquez-Torres, A. (2011). Nitric oxide and salmonella pathogenesis. *Front. Microbiol.* 2:84. doi: 10.3389/fmicb.2011.00084
- Henderson, B., Allan, E., and Coates, A. R. M. (2006). Stress wars: the direct role of host and bacterial molecular chaperones in bacterial infection. *Infect. Immun.* 74, 3693–3706. doi: 10.1128/IAI.01882-05
- Ho, N. K., Hawley, S. P., Ossa, J. C., Mathieu, O., Tompkins, T. A., Johnson-Henry, K. C., et al. (2013a). Immune signalling responses in intestinal epithelial cells exposed to pathogenic *Escherichia coli* and lactic acid-producing probiotics. *Benef. Microbes* 4, 195–209. doi: 10.3920/BM2012.0038
- Ho, N. K., Henry, A. C., Johnson-Henry, K., and Sherman, P. M. (2013b). Pathogenicity, host responses and implications for management of enterohemorrhagic *Escherichia coli* O157:H7 infection. *Can. J. Gastroenterol.* 27, 281–285.
- Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35, W585–W587. doi: 10.1093/nar/gkm259
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13. doi: 10.1093/nar/gkn923
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211
- Jia, B., Cheong, G.-W., and Zhang, S. (2013). Multifunctional enzymes in archaea: promiscuity and moonlight. *Extremophiles* 17, 193–203. doi: 10.1007/s00792-012-0509-1
- Jung, I. L., and Kim, I. G. (2003). Transcription of *ahpC*, *katG*, and *katE* genes in *Escherichia coli* is regulated by polyamines: polyamine-deficient mutant sensitive to H₂O₂-induced oxidative damage. *Biochem. Biophys. Res. Commun.* 301, 915–922. doi: 10.1016/S0006-291X(03)00064-0
- Junjie, Q., Ruiqiang, L., Jeroen, R., Manimozhayan, A., Kristoffer, S. B., Chaysavanh, M., et al. (2010). A human gut microbial gene catalog established by metagenomic sequencing. *Nature* 464, 59–65. doi: 10.1038/nature08821
- Justice, S. S., Hunstad, D. A., Harper, J. R., Duguay, A. R., Pinkner, J. S., Bann, J., et al. (2005). Periplasmic peptidyl prolyl cis-trans isomerases are not essential for viability, but SurA is required for pilus biogenesis in *Escherichia coli*. *J. Bacteriol.* 187, 7680–7686. doi: 10.1128/JB.187.22.7680-7686.2005
- Karch, H., Tarr, P. I., and Bielaszewska, M. (2005). Enterohaemorrhagic *Escherichia coli* in human medicine. *Int. J. Med. Microbiol.* 295, 405–418. doi: 10.1016/j.ijmm.2005.06.009
- Karmali, M. A. (1989). Infection by verocytotoxin-producing *Escherichia coli*. *Clin. Microbiol. Rev.* 2, 15–38. doi: 10.1128/CMR.2.1.15
- Kim, C., Song, S., and Park, C. (1997). The D-allose operon of *Escherichia coli* K-12. *J. Bacteriol.* 179, 7631–7637. doi: 10.1128/jb.179.24.7631-7637.1997
- Kocianova, S., Vuong, C., Yao, Y., Voyich, J. M., Fischer, E. R., DeLeo, F. R., et al. (2005). Key role of poly-gamma-DL-glutamic acid in immune evasion and virulence of *Staphylococcus epidermidis*. *J. Clin. Invest.* 115, 688–694. doi: 10.1172/JCI200523523
- Kuntal, B. K., Dutta, A., and Mande, S. S. (2016). CompNet: a GUI based tool for comparison of multiple biological interaction networks. *BMC Bioinformatics* 17:185. doi: 10.1186/s12859-016-1013-x
- Lafont, F., Tran Van Nhieu, G., Hanada, K., Sansonetti, P., and van der Goot, F. G. (2002). Initial steps of Shigella infection depend on the cholesterol/sphingolipid raft-mediated CD44-IpaB interaction. *EMBO J.* 21, 4449–4457. doi: 10.1093/emboj/cdf457
- Lee, J.-H., Kim, Y.-G., Ryu, S. Y., Cho, M. H., and Lee, J. (2014). Ginkgolic acids and *Ginkgo biloba* extract inhibit *Escherichia coli* O157:H7 and *Staphylococcus aureus* biofilm formation. *Int. J. Food Microbiol.* 174, 47–55. doi: 10.1016/j.ijfoodmicro.2013.12.030
- Li, Y., Frey, E., Mackenzie, A. M., and Finlay, B. B. (2000). Human response to *Escherichia coli* O157:H7 infection: antibodies to secreted virulence factors. *Infect. Immun.* 68, 5090–5095. doi: 10.1128/IAI.68.9.5090-5095.2000
- Liang, W.-J., Wilson, K. J., Xie, H., Knol, J., Suzuki, S., Rutherford, N. G., et al. (2005). The *gusBC* genes of *Escherichia coli* encode a glucuronide transport system. *J. Bacteriol.* 187, 2377–2385. doi: 10.1128/JB.187.7.2377-2385.2005
- Lowe, R. M. S., Baines, D., Selinger, L. B., Thomas, J. E., McAllister, T. A., and Sharma, R. (2009). *Escherichia coli* O157:H7 strain origin, lineage, and Shiga toxin 2 expression affect colonization of cattle. *Appl. Environ. Microbiol.* 75, 5074–5081. doi: 10.1128/AEM.00391-09
- Ma, Z., Richard, H., Tucker, D. L., Conway, T., and Foster, J. W. (2002). Collaborative regulation of *Escherichia coli* glutamate-dependent acid resistance by two AraC-like regulators, GadX and GadW (YhiW). *J. Bacteriol.* 184, 7001–7012. doi: 10.1128/JB.184.24.7001-7012.2002
- Mademidis, A., and Köster, W. (1998). Transport activity of FhuA, FhuC, FhuD, and FhuB derivatives in a system free of polar effects, and stoichiometry of components involved in ferrichrome uptake. *Mol. Gen. Genet.* 258, 156–165. doi: 10.1007/s004380050718
- Malladi, V., Shankar, B., Williams, P. H., and Balakrishnan, A. (2004). Enteropathogenic *Escherichia coli* outer membrane proteins induce changes in cadherin junctions of Caco-2 cells through activation of PKC α . *Microbes Infect.* 6, 38–50. doi: 10.1016/j.micinf.2003.09.022
- Mariappan, V., Vellamy, K. M., Thimma, J., Hashim, O. H., and Vadivelu, J. (2013). Infection of *Burkholderia cepacia* induces homeostatic responses in the host for their prolonged survival: the microarray perspective. *PLoS ONE* 8:e77418. doi: 10.1371/journal.pone.0077418
- Maurer, L. M., Yohannes, E., Bondurant, S. S., Radmacher, M., and Slonczewski, J. L. (2005). pH regulates genes for flagellar motility, catabolism, and oxidative stress in *Escherichia coli* K-12. *J. Bacteriol.* 187, 304–319. doi: 10.1128/JB.187.1.304-319.2005
- McLean, S., Bowman, L. A. H., Sanguinetti, G., Read, R. C., and Poole, R. K. (2010). Peroxynitrite toxicity in *Escherichia coli* K12 elicits expression of oxidative stress responses and protein nitration and nitrosylation. *J. Biol. Chem.* 285, 20724–20731. doi: 10.1074/jbc.M109.085506
- Motley, S. T., Morrow, B. J., Liu, X., Dodge, I. L., Vitiello, A., Ward, C. K., et al. (2004). Simultaneous analysis of host and pathogen interactions during an *in vivo* infection reveals local induction of host acute phase response proteins, a novel bacterial stress response, and evidence of a host-imposed metal ion limited environment. *Cell. Microbiol.* 6, 849–865. doi: 10.1111/j.1462-5822.2004.00407.x
- Murooka, Y., Higashiura, T., and Harada, T. (1978). Genetic mapping of tyramine oxidase and arylsulfatase genes and their regulation in intergeneric hybrids of enteric bacteria. *J. Bacteriol.* 136, 714–722.
- Neuber, O., Jarosch, E., Volkwein, C., Walter, J., and Sommer, T. (2005). Ubx2 links the Cdc48 complex to ER-associated protein degradation. *Nat. Cell Biol.* 7, 993–998. doi: 10.1038/ncb1298
- Rossez, Y., Holmes, A., Wolfson, E. B., Gally, D. L., Mahajan, A., Pedersen, H. L., et al. (2014). Flagella interact with ionic plant lipids to mediate adherence of pathogenic *Escherichia coli* to fresh produce plants. *Environ. Microbiol.* 16, 2181–2195. doi: 10.1111/1462-2920.12315

- Rossez, Y., Wolfson, E. B., Holmes, A., Gally, D. L., and Holden, N. J. (2015). Bacterial flagella: twist and stick, or dodge across the kingdoms. *PLoS Pathog.* 11:e1004483. doi: 10.1371/journal.ppat.1004483
- Ruggieri, M. R., Hanno, P. M., and Levin, R. M. (1985). Mannose inhibition of *Escherichia coli* adherence to urinary bladder epithelium: comparison with yeast agglutination. *Urol. Res.* 13, 79–84. doi: 10.1007/BF00261571
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Srinivasan, S., and Venkatesh, K. V. (2014). Steady state analysis of the genetic regulatory network incorporating underlying molecular mechanisms for anaerobic metabolism in *Escherichia coli*. *Mol. Biosyst.* 10, 562–575. doi: 10.1039/c3mb70483a
- Tildesley, M. J., Gally, D. L., McNeilly, T. N., Low, J. C., Mahajan, A., and Savill, N. J. (2012). Insights into mucosal innate responses to *Escherichia coli* O157:H7 colonization of cattle by mathematical modelling of excretion dynamics. *J. R. Soc. Interface* 9, 518–527. doi: 10.1098/rsif.2011.0293
- Torres, A. G., Li, Y., Tutt, C. B., Xin, L., Eaves-Pyles, T., and Soong, L. (2006). Outer membrane protein A of *Escherichia coli* O157:H7 stimulates dendritic cell activation. *Infect. Immun.* 74, 2676–2685. doi: 10.1128/IAI.74.5.2676-2685.2006
- Tyagi, N., Krishnadev, O., and Srinivasan, N. (2009). Prediction of protein-protein interactions between *Helicobacter pylori* and a human host. *Mol. Biosyst.* 5, 1630–1635. doi: 10.1039/b906543c
- Ukena, S. N., Westendorf, A. M., Hansen, W., Rohde, M., Geffers, R., Coldewey, S., et al. (2005). The host response to the probiotic *Escherichia coli* strain Nissle 1917: specific up-regulation of the proinflammatory chemokine MCP-1. *BMC Med. Genet.* 6:43. doi: 10.1186/1471-2350-6-43
- Wallrodt, I., Jelsbak, L., Thorndahl, L., Thomsen, L. E., Lemire, S., and Olsen, J. E. (2013). The putative thiosulfate sulfurtransferases PspE and GlpE contribute to virulence of *Salmonella* Typhimurium in the mouse model of systemic disease. *PLoS ONE* 8:e70829. doi: 10.1371/journal.pone.0070829
- Wang, S., Deng, K., Zaremba, S., Deng, X., Lin, C., Wang, Q., et al. (2009). Transcriptomic response of *Escherichia coli* O157:H7 to oxidative stress. *Appl. Environ. Microbiol.* 75, 6110–6123. doi: 10.1128/AEM.00914-09
- Wu, R., Xie, H., Cao, Z., and Mo, Y. (2008). Combined quantum mechanics/molecular mechanics study on the reversible isomerization of glucose and fructose catalyzed by *Pyrococcus furiosus* phosphoglucose isomerase. *J. Am. Chem. Soc.* 130, 7022–7031. doi: 10.1021/ja710633c
- Wuchty, S. (2011). Computational prediction of host-parasite protein interactions between *P. falciparum* and *H. sapiens*. *PLoS ONE* 6:e26960. doi: 10.1371/journal.pone.0026960
- Yu, H., Luscombe, N. M., Lu, H. X., Zhu, X., Xia, Y., Han, J.-D. J., et al. (2004). Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. *Genome Res.* 14, 1107–1118. doi: 10.1101/gr.1774904
- Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., et al. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26, 1608–1615. doi: 10.1093/bioinformatics/btq249
- Zhang, Y., Laing, C., Steele, M., Ziebell, K., Johnson, R., Benson, A. K., et al. (2007). Genome evolution in major *Escherichia coli* O157:H7 lineages. *BMC Genomics* 8:121. doi: 10.1186/1471-2164-8-121
- Zhang, Y., Laing, C., Zhang, Z., Hallewell, J., You, C., Ziebell, K., et al. (2010). Lineage and host source are both correlated with levels of Shiga toxin 2 production by *Escherichia coli* O157:H7 strains. *Appl. Environ. Microbiol.* 76, 474–482. doi: 10.1128/AEM.01288-09
- Zhou, H., Gao, S., Nguyen, N. N., Fan, M., Jin, J., Liu, B., et al. (2014). Stringent homology-based prediction of *H. sapiens*-*M. tuberculosis* H37Rv protein-protein interactions. *Biol. Direct* 9:5. doi: 10.1186/1745-6150-9-5

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

Copyright © 2017 Bose, Venkatesh and Mande. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.