Genome Sequencing-Based Evolutionary Prediction of Glycan-Adhering Lectins and Experimental Evaluation of a Lectin (FimH) Inhibitor Using an Enterohemorrhagic Escherichia coli Strain Isolated in Korea

Junyoung Park
Korea National Institute of Health
Cheorl-Ho Kim
Sung Kyunkwan University and Samsung Advanced Institute for Health Science and Technology (SAIHST)
Seung-Hak Cho (skcho38@korea.kr)
Korea National Institute of Health

Article

Keywords:

Posted Date: April 14th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1548018/v1

Abstract

We analyzed the evolutionary characteristics of the predicted lectin-like adhesins to construct a lectin-glycan interaction network (LGI). A total of 20,603 assembled Esherichia. coli sequences were retrieved from NCBI, including 2,240 strains with Shiga toxin in their genome. A strain isolated from a Korean patient with diarrhea, EHEC NCCP14539, was also sequenced. The outer-membrane embedded proteins which could be adhesins were predicted from the annotated genes of E. coli strains. In addition, we performed a phylogenetic analysis using the fimH gene, a well-known adhesin of enterohemorrhagic E. coli (EHEC). Multiple sequence alignment of the 2,204 E. coli strains yielded an evolutionary tree. Multi-locus sequence typing-based phylogenetic analysis revealed that NCCP14539 is the ancestor of 70 EHEC strains. An experiment was conducted to find an FimH inhibitor using NCCP14539 strain. Using a T7 phage display method, FimH was selected as an adherent lectin from the GM1- and Gb3- attached lectins. Through simulation docking, we then confirmed that Gb3 has a stronger binding ability to FimH and inferred the binding site sequence. Thus, our results provide evidence that Gb3-derived peptides can act as novel agents to block FimH binding and prevent infection by EHEC.

Introduction

A class of virulence proteins, called lectin-like virulence proteins, is responsible for the pathogenesis of bacterial infections, and it may be exploited as a therapeutic target or vaccine components [1]. Bacterial adhesins are lectin proteins that have the ability to adhere to host cells and they have a variety of structural topologies [2]. Pili and fimbriae are some of the structures found in bacteria. These bind to host cell surface receptor proteins and participate in a variety of biological processes, including cross-membrane tracking, invasion, and migration [3]. They can be toxic to the host, causing inflammation [4]. Mannose supplementation or receptor inhibition may affect the adhesin–receptor interaction in certain adhesins, such as mannose, which is critical in immunological activation. The PilA glycoprotein from Acinetobacter baumannii binds to host cell selectins and carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) [5]. Other lectin-like proteins include *Escherichia coli*'s surface antigen 20 (CS20) and fimbriae protein SfaS; Hemophilus influenzae's surface-adhesin protein E; Neisseria meningitidis's autotransporter adhesin; and Salmonella enterica serovar Enteritidis's ShdA, MisL, Sad, and BapA [6]. These proteins adhere to the host gut by the binding of lectin-like adhesins to host cell receptors that comprise glycans, resulting in lectin-glycan interactions (LGIs) [7].

E. coli is rod-shaped gram-negative bacterium that is often found in the lower intestines of warm-blooded organisms, including humans [8]. Most *E. coli* serotypes are non-pathogenic, but some serotypes can cause food poisoning, such as Enterohemorrhagic *Escherichia coli* (EHEC). One of the major serotypes of this group is O157:H7. Infection with this type of pathogenic *E. coli* can lead to hemorrhagic colitis or hemolytic uremic syndrome (HUS) [9, 10]. Since the first confirmation of O157:H7 as a human pathogen in 1982, sporadic outbreaks of it infection have been reported in various regions of the world. There has been a long-term epidemic in Japan, and this strain has been recognized as an epidemiologically important infectious pathogen [11]. The Shiga-like toxins Stx1 and Stx2 are expressed in *E. coli* O157:H7. The genes encoding these toxins may be horizontally transmitted to *E. coli* or other *Enterobacteriaceae* species through prophage [12], enabling the transformation of non- Shiga-like toxin-generationg strains into Shiga-like toxin producing strains [13]. Shiga toxins produced by Shiga toxin-producing *E. coli* (STEC) are mostly directed at capillary endothelial cells. Shiga toxins specifically target the globotriaosylceramide receptor on cells, and bacteria enter cells through receptor-mediated endocytosis [14]. Shiga toxins prevent protein synthesis by cleaving an adenine base from infected cells' ribosomes [15]. As demonstrated in HUS [14], this obstruction may result in renal failure. Despite the prevalence of serotype STEC O157 strains, we identified a novel STEC O26 strain, EHEC NCCP14539. We sequenced its genome to uncover novel genes and to gain a better understanding of the strain's evolution.

The majority of gram-negative bacteria produce adhesins, and in many insstances, the adhesin's activity is mediated by a tiny component protein located at the tip of each fimbriae. For example, the bacterial adhesin FimH is responsible for the adhesion of *E. coli* to host intestinal cells via the activation of their glycan receptors, such as CD48 and TLR4, or via binding to mannose residues. Various attempts have been made to produce anti-adhesion vaccines, including those based on anti-FimH antibodies. Passive immunization with anti-FimH antibodies has been demonstrated to significantly reduce uropathogenic *E. coli* colonization in animal models [16]. On the contrary,the lectin-like adhesins of *E. coli*, are not currently catalogued, making their subsequent use in vaccine development more difficult. As a result, further exploration of *E. coli* adhesins at the genome-wide level is desirable.

Thus, in this study, we used whole-genome sequencing to elucidate the genetic origins and evolution of NCCP14539, which was isolated from a patient with diarrhea. Additionally, we compared the genomes of NCCP14539 and a reference strain, EDL933, to examine their evolution and phylogenetic relationship [17]. Furthermore, we attempted to isolate FimH binding inhibitors using T7 phage display, docking simulation, and a binding test, in order to facilitate the discovery of therapeutics for the prevention of EHEC infection during the initial stage of the illness.

Results

Genomic features of NCCP14539. A total of 325,481,384 (39,050,715,080 bp) paired-end reads were generated using IonTorrent. Using the PacBio RS II platform, 52,853 (482,609,715 bp) raw reads were produced. The complete genome of NCCP14539 consists of a 5,315,402 bp circular chromosome. Rapid Annotation using Subsystem Technology (RAST) analysis revealed 4,555 putative open reading frames (ORFs) and 31 RNA genes, of which 4,181 (80.6%) could be functionally annotated (Fig. 1). Among the subsystems, the central glucose metabolism (439 ORFs) and amino acids and derivatives (342 ORFs) subsystems were considerably more abundant (17.1%). We speculate that NCCP14539 has evolved systems that can utilize various carbohydrates in addition to protein in order to adapt to extreme environment in the hosts. A large number of ORFs was also associated with the “Amino acids and derivatives” (342 ORFs), “Membrane transport” (114 ORFs) and “cofactors, vitamins, prosthetic groups, and pigments” (174 ORFs) subsystems. Although NCCP14539 belongs to serotype O26, it has more virulence factors (213) than EDL933. As a result, we were interested in unveiling novel virulence factors in the NCCP14539 genome. Table 2 summarizes the traits discovered using sequence analysis, and they include a variety of pilus and fimbriae genes and their related operons.

NCCP14539 only generates the Shiga toxin Stx1 (stx1A and stx1B). NCCP14539 possesses nine distinct virulence genes, as determined by comparison with *E. coli* O157:H7 str. EDL933. NCCP14539 has 26 virulence genes, 17 of which have previously been identified in two the other strains.

Evolutionary analysis of NCCP14539 with other EHEC strains. The phylogenetic comparison of gene candidates predicted by seed [18] revealed Shigella flexneri 2a str. 2457T as the closest neighbor of NCCP14539 (score 524). To determine the complete evolutionary history of NCCP14539, we conducted a multiple sequence alignment analysis of 2,204 STEC strains, including NCCP14539 (Fig. 2, Supplementary Materials 1: Table S1) and S. flexneri 2a str. 2457T (Fig. 2, Supplementary Materials 1: Table S1). Multi-locus sequence typing (MLST)-based phylogenetic analysis revealed that NCCP14539 did not cluster with *E. coli* O157:H7 str. EDL933 nor S. flexneri 2a str. 2457T. NCCP14539 was an ancestor of 70 EHEC strains rather than a sibling. Meanwhile, two NCCP strains, NCCP15736 and NCCP15737, previously reported by us were near EDL933, but NCCP15738 and NCCP15739 were distant from EDL933 in the phylogenetic tree (Fig. 2). A comparison of NCCP14539 genes with those of the reference strain *E. coli* O157:H7 str. EDL933 revealed that although the majority of NCCP14539's functional genes were preserved in the reference strain, 236 genes were unique (Supplementary Materials 2: Table S2). The genes encoding phage-associated proteins, a two-component signal transduction system, conjugation, the flagellum, nucleotide-binding proteins, and metal ion binding proteins were all unique in NCCP14539 indicating phenotypic variations resulting from environmental adaptation. We speculate that horizontal gene transfer (HGT) in NCCP14539 is the major machinery for transferring virulence factors.

Gene-based subtyping of EHEC strains in relation to LGI. Owing to the vast diversity of *E. coli*, reliance on gene-based subtyping may result in the identification of new groups of closely related isolates as a result of gene variation or links among unrelated isolates due to recombination. Therefore, we used a gene-based technique to classify the genomes into lineages of closely related isolates. For the gene-based subtyping, we selected a gene, fimH, as the gene encodes a major adhesin in EHEC strains. Based on the gene-based subtyping described above, the grouping produced two major lineages. NCCP14539 belongs to the large clade (Fig. 3) and this suggests that the strain has a general mechanism of lectin-glycan interaction. We predicted putative adhesins-coding genes in the NCCP14539 genome, compared them with those in EDL933 and other NCCP strains, and found that NCCP14539 has less putative adhesins than the other strains. Table 3 summarizes the potential adhesins found using the e-Membranome pipeline, which also contains a variety of pilus and fimbriae genes and their related operons. When potential adhesins were selected, they were examined using the e-membranome evaluation index calculation formula [27, 43].

Phage-displayed peptide library to find a lectin receptor: FimH in EHEC NCCP14539 strain. In this method, a phage infects bacteria and is exhibited on a bead covered with glycan, and the DNA of the adherent phage is sequenced to reveal the encoded proteins, thus detecting bacterial lectin. cDNA of each strain was synthesized, and a phage library was created. Immunoscreening processes were carried out in accordance with industry standards. Approximately 500 plaques were produced from each biopanning, and then, they were placed on separate Hybond TM-N + nylon membranes. A vast number of phage clones were analyzed in this manner, and each clone was created from a distinct virus. The plaque lift assay revealed lectin factors linked to Gb3 and GM1a glycan (Fig. 4A). Furthermore, this method enabled us to decipher the structure of the pathogen's lectin and to identify the site of attachment to the host's receptor. Docking simulations with FimH, a lectin candidate that binds to Gb3 and GM1a glycans, were used to determine binding affinity. We chose residues with a binding affinity of -3 kcal/mol or less for the FimH lectin residue interacting with the ligand, and − 3 kcal/mol or less for other ligands, to locate the lectin-binding domain (LBD) of FimH. In addition, the binding strengths of GM1, Gb3, and FimH lectins were higher for Gb3 than for GM1a. Three LBDs at the anticipated binding locations were discovered as a result (Fig. 4B and C). Therefore, FimH was analyzed as a possible lectin receptor.

Comparison of the adhesion affinity of pathogenic enteric bacteria EHEC NCCP14539 strain. From our analysis, three LBDs were identified. The peptide candidate sequences for the binding site are mannose-6-phosphonate-conjugated CGTVLTRNETHATYS, CQCKQDFNITDISLL, and CYATPSSNATDPLKY, which were named P1, P2, and P3, respectively. An adhesion test with FimH lectin of the separating strain was conducted. P2 was found to have the strongest binding strength. Compared with the control, P1, P2, and P3 showed significant variations in their capacity to bind to FimH (Fig. 5). We speculated that the Gb3 glycan-like peptide may bind to FimH and affect its binding to host glycans.

Discussion

Both non-experimental and experimental methodologies are required to predict lectin candidate proteins in pathogenic bacteria. Genome-based screening is a non-experimental technique to identify lectins from pathogenic bacteria [6]. EHEC strains are recognized as epidemiologically important infectious pathogens as sporadic outbreaks of their infections can result in an epidemic [19, 20]. Therefore, it is important to perform genomic sequencing and investigate their evolution [21]. To characterize the evolution of EHEC, we performed complete genome sequencing of the newly identified STEC O26 strain NCCP14539 and conducted comparative genomic analysis. The genome analysis showed that NCCP14539 possesses more metabolic systems, membrane transport subsystems, and virulence factors than EDL933, although NCCP14539 also belongs to serotype O26. Moreover, the analysis of gene-based subtyping using the fimH gene suggested that NCCP14539 belongs to the large group of EDL933. From these results, it can be assumed that NCCP14539 has a general mechanism and machinery of lectin-glycan interaction of EHEC strains. Therefore, we used NCCP14539 for inhibitor evaluation against FimH.

FimH, which is a mannose-specific adhesin that mediates shear-enhanced bacterial adherence and is found at the tip of type 1 fimbriae of E. coli, plays an important role in uropathogenic E. coli (UPEC) for the binding and colonization of urothelial cells (uroplakins), as well as UPEC invasion of bladder epithelial cells [22]. FimH facilitates bacterial attachment to urothelial cells, which is the first stage in the infection process [23]. Itis made up of two domains, a pilin domain associated with the fimbria and a mannose-binding lectin domain, with the binding pocket on the other side of the interdomain interface [24]. The anti-FimH activity of several mannose derivatives has been shown and are consequently regarded to be potential therapeutic agents for the treatment of urinary tract infections [25, 26]. However, in the case of EHEC and ETEC, the binding to FimH is not well understood. Thus, it is important to investigate the role of FimH in EHECs.

Several studies have elucidated the interaction between pathogen lectins and host glycans during the early stages of infection [6, 7, 27]. We established an experimental strategy and method to characterize the systemic LGI of pathogenic intestinal bacteria in the gastrointestinal tract. Thus, we conducted various

experiments to analyze FimH inhibitors using NCCP14539, a strain isolated from a patient with diarrheal, as field isolates can be more useful for the discovery of new medical agents than reference strains such as EDL933. We used the T7 phage display method to identify binding peptides or proteins to glycans [28–30]. Plaques with a high concentration of GM1a and Gb3 were found to be associated with intestinal epithelial cells; thus, novel FimH-related lectins were discovered.

Additionally, we demonstrated that Gb3 had a better binding capability to FimH than GM1a and used simulated docking to determine the binding site sequence. Docking modeling was used to predict the structures of sugar chains that may bind to lectin ligands. These protein-ligand and sugar chain-lectin ligand docking approaches were used to estimate the binding structure of ligands. Additionally, these strategies were used to identify and optimize binding sites and candidate materials [6, 7, 27].

The array technology, which is used to assess the specificity of the pathogen lectin for the host glycan, can be used to evaluate hundreds of samples concurrently; moreover, itrequires only a small quantity of each sample [31]. By using qualitative or quantitative assessments of the specificity of the pathogen lectin for the host glycan, this approach may be utilized to identify a likely host glycoprotein that binds to an adhesin of a pathogen [32]. Here, we sought to determine whether the adhesin was bound using the three Gb3 attachment peptide candidate groups P1, P2, and P3 and found that P2 has the highest binding affinity. These results suggest that the peptide may act as an important inhibitor restraining the adherence of EHEC to epithelial cells of the human gastrointestinal tract during the initial stage of EHEC infection (Fig. 6). However, more experiments for the confirmation of binding capacity using cell lines or animal models are needed. The protein sequences and structures of pathogen lectins may aid the development of a vaccine or antibiotic substitute for therapeutic usage.

Conclusions

We examined how pathogen lectins interact with host glycans early in infection. Reducing the treatment costs for intestinal infections will benefit both public health and the economy. We recommend further research on lectins and glycans to lay a foundation for developing therapies for gastrointestinal infections.

Methods

Strain, isolation, and serotyping. A fecal sample from a patient with diarrhea was plated on MacConkey agar directly or after enrichment in trypticase soy broth containing vancomycin (Sigma Chemicals Co., St. Louis, MO). Candidate colonies were then plated on trypticase soy agar medium and biochemically characterized using the API20E system (Biomerieux, Marcy l'Etoile, France). For O-antigen determination, we used the method described by Guinee et al [33], using all available O (O1 to O181) antisera. Antisera were absorbed with the corresponding cross-reacting antigens to remove non-specific agglutinins. The O antisera were produced at Laboratorio de Referencia de *E. coli* (Lugo, Spain [http://www.lugo.usc.es/ecoli]). This research was approved by the Research Ethics Committee of the Korea Centers for Disease Control and Prevention, and written informed consent was obtained from the patient. The isolated strain was deposited at the National Culture Collection for Pathogens (NCCP) at the Korea National Institute of Health under the accession number NCCP14539. EHEC O157:H7 str. EDL933 was used as the reference strain. In addition, we included previously published EHEC strains that were isolated from Korea: NCCP15736 [34], NCCP15737 [34], NCCP15738 [35] and NCCP15739 [36] for comparative genomic analysis (Table 1).

Library preparation and whole-genome sequencing. The genomic DNA of NCCP14539 was extracted from a culture of candidate colonies using a Wizard Genomic DNA purification kit (Promega, USA) according to the manufacturer's instructions. Genomic DNA yield, purity, and concentration were evaluated using 0.8% agarose gel electrophoresis and a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific, USA). Three kinds of libraries were constructed for three platforms (DNALink: Seoul, Republic of Korea). First, a mate-paired library with 1 to 10 kb insert size was constructed for the IonTorrent platform. Secondly, PacBio RS II libraries were constructed and sequenced with 8 to 20 kb insert sizes. Third, a fosmid library with 30 kb average insert size (CopyControl Fosmid library production kit, Epicenter, Madison, WI) was constructed and used as a template for the construction of the physical map. 325,481,384 reads (39,050,715,080 bp) of raw reads with IonTorrent and 52,853 reads (482,609,715 bp) of raw reads with PacBio RS II were produced.

Genome assembly and annotation. A hybrid assembly was generated using SPAdes assembler (version 3.1) and confirmed with a FES (Fosmid End Sequence) map. The sequence gaps between contigs were filled with Sanger sequencing data after PCR amplification. The final assembly was corrected with proovread (version 2.12) [37], as the number of frameshifted genes was greater than 5% of the called genes. The genome was determined to be composed of a 5,315,402 bp circular chromosome. To predict and annotate ORFs, we used the RAST version 4.0 [38] server pipeline. To identify the virulence factor genes in NCCP14539, we performed a Basic Local Alignment Search Tool (BLAST) search of NCCP14539 ORFs against the virulence factor genes listed in VFDB [39], with a cutoff e-value of 1e-5.

Phylogenetic analysis and comparative genomic analysis. A total of 20,603 *E. coli* genomes were downloaded from the National Center for Biotechnology Information (NCBI) and curated to create a final collection of 2,204 EHEC genomes with Shiga-toxin genes. To infer the evolutionary history of NCCP14539, we performed a multiple sequence alignment of the MLST genes using MEGA X (version 10.0.5). A MUSCLE algorithm was used for multiple sequence sorting, and the Unweighted Pair Group Method with Arithmetic Mean method was used as the clustering algorithm. Multiple sequence alignment results were saved in Mega format. Maximum-Parsimony was used as a statistical test method for estimating phylogenetic trees. The phylogenetic tree test was performed using the bootstrap method, with 1,000 iterations. The substitution model used maximum composite likelihood and included both transitions and transversions. Calculation results were saved in nexus format. The tree was visualized with FigTree (version 1.3.1) (http://tree.bio.ed.ac.uk/software/figtree/). In order to exclude the effect of HGT in our phylogenetic analysis, we used the multi-locus sequence analysis method [40, 41]. Seven housekeeping genes (adk, fumC, gyrB, icd, mdh, purA, and recA) from 2,204 *E. coli* strains were retrieved and concatenated. A phylogenetic tree of MLST genes was created using the method employed for MLST-based phylogenetic analysis [42]. From a comparative genomic study using SEED Viewer version 2.0, we identified a syntenic region that aligned with the reference genome. Unaligned genes were defined as unique genes of NCCP14539.

Gene-based subtyping of EHEC strains in relation to LGI. During infection, EHEC strains attach to the host intestinal cell by the binding of adhesins to receptors of the host, such as glycans. Bacterial adhesins such as lectin are potential therapeutic targets and can contribute to an understanding of bacterial evolution. In our previous study, we constructed the LGI network of EHEC [6] and developed an e-Membranome database [27]. Using an e-Membranome pipeline, we predicted putative adhesins from EHEC genomes and deposited them onto a public database. Among the putative adhesins, we selected the fimH gene and performed gene-based subtyping of EHEC strains. The nucleotide sequences of fimH gene were retrieved from 1,482 EHEC strains, including NCCP14539. We performed a phylogenetic analysis of the fimH gene using MEGA X (version 10.0.5), as described above.

T7 phage display. The biopanning and phage propagation procedures were carried out according to the manufacturer's instructions. We performed differential biopanning with negative and positive selection using pre-immune rabbit serum and rabbit polyclonal antibody as indicated by the manufacturer (TB178 T7Select® System; Novagen). Twenty-five liters of protein G + agarose beads were washed in phosphate-buffered saline (PBS) and blocked in 1% bovine serum albumin for 1 hour (BSA). After two hours of incubation with pre-immune rabbit serum at a dilution of 1:20, the beads were washed three times with PBS and then incubated for another two hours with the T7 phage display human brain cDNA library. This subtractive biopanning approach is required to eliminate proteins that react with pre-immune IgGs. The supernatant T7 phage cDNA library was then treated with polyclonal antibodies immobilized on protein G + agarose beads and incubated overnight at 4°C. After three washes with 1 PBS, the bound T7 phage cDNA library was eluted with 1% sodium dodecyl sulfate. The eluant was amplified using *E. coli* strain BLT5616 for the subsequent cycle of biopanning. After four rounds of biopanning, the selected phage library was used for immunoscreening.

Docking simulation. Using SWISS-MODEL, a homology modeling approach, the three-dimensional (3D) structure of FimH (PDB ID: 6GTY and Sequence :MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGAAYGGVLSSFSGTVKYNGSSYPFPTTSETPI was produced using GM1(PDB ID: 4ZH1)- and Gb3(PDB ID: 6F4C) 3D structure. The docking simulation was conducted using the 3D structure of GM1 and Gb3. The Chem-office application (http://www.cambridgesoft.com, version: 7.0) was also used to reduce energy use. Autodock Vina 1.1.2 was used for the docking simulations. Possible hydrogen bonds and hydrophobic interactions were identified utilizing HBPLUS and non-bonded contact parameters as default settings, in a LigPlot based on the findings of docking simulation.

Plaque lift assay. EHEC bacteria in log phase growth were infected with various dilutions of phage and grown on agar plates. About 1000 well-dispersed individual phage clones from agar plates were transferred to nitrocellulose membranes using the plaque-lift technique. A nitrocellulose membrane (82-mm diameter HATF filter, Millipore Corp.) was overlaid onto the agar and incubated at room temperature for 10–15 min. Without disrupting the agar and plaques, the nitrocellulose membrane was carefully lifted. The immunostaining procedure included an initial incubation with the binding glycan (GM1 and Gb3, respectively)-EHEC for 1–2 h. The membranes were then rinsed, followed by the application of detecting reagents. The detecting reagents include horseradish peroxidase-conjugated goat anti-mouse IgG, followed by color development using 3, 3-diaminobenzidine (DAB) and 0.01% H2O2.

Solid Phase Peptide Synthesis. The initial step in synthesizing peptides on a resin was to connect the C-terminal amino acid to the resin. A transient protective group protects the alpha amino group and the reactive side chains from polymerization. The resin was next filtered and washed to eliminate byproducts and excess reagents. The N-alpha protective group was then removed, and the resin was washed again to eliminate byproducts and excess reagent. Then, the next amino acid was added until the peptide sequence was complete. It was then rinsed to remove the protective groups and the peptide was released from the resin. As in the above procedure, the peptide manufacturing synthesis proceeded with the same protocol for the mannose-6-phosphonate conjugated P1, P2, and P3. The Gb3-like peptides, P1 (CGTVLTRNETHATYS), P2 (CQCKQDFNITDISLL), and P3 (CYATPSSNATDPLKY) were similarly synthesized.

Peptide binding assays. The mannose-6-phosphonate-conjugated P1, P2, and P3 peptides were coated to each well of the plate at a concentration of 25 ng per well. After blocking, each well was filled with EHEC for 2 hours at room temperature and washed five times with 1% BSA. Next, the plate was incubated at 4°C for 15–16 hours. After another wash, the plate was stained with fluorescein isothiocyanate-conjugated staining reagent at a 1: 10,000 dilutions. After 2 hours at room temperature, the plates were washed, followed by fluorescence detection using a micro reader.

Statistical analysis. All experiments were carried out at least three times, and representative results are shown. The outcomes of the data analysis were statistically analyzed using the comparison-based one-way analysis of variance (ANOVA), which was then followed by a post hoc Bonferoni test to determine significance. Differences were considered statistically significant when their p-values were less than 0.05. *p indicates < 0.05 and **p < 0.01. The differences between the two figures are indicated in the figure legends.

Declarations

Data Availability Statement

Nucleotide sequence accession numbers: This complete genome project has been deposited in DDBJ/EMBL/GenBank under the accession number GCA_009873415.1.

Acknowledgments

This work was supported by a grant from the Korea National Institute of Health (NIH 4800-6637-300 to S.H.C.).

Funding

This work was supported by a grant from the Korea National Institute of Health (NIH 4800-6637-300 to S.H.C.).

Author Contributions

Conceptualization, S.-H.C. and J.P.; writing—original draft preparation, S.-H.C and J.P.; writing—review and editing, J.P. and C.-H.K.; visualization, J.P.; supervision, C.-H.K. and S.-H.C.; project administration, S.-H.C.; funding acquisition, S.-H.C. All authors have read and agreed to the published version of the manuscript.

Ethics approval and consent to participate

This study was approved by the Research Ethics Committee of the Korean Centers for Disease Control and Prevention and written informed consent was obtained from the patient.

Competing interests

The authors declare no conflict of interests.

Additional information

Supplementary Materials: Table. S1. Bacterial strain list used for phylogenetic analysis. Table. S2. Annotated genes of STEC O26 NCCP14539 using.

References

1. Veronika Tchesnokova, Pavel Aprikian, Dagmara Kisiela. Type 1 fimbrial adhesin FimH elicits an immune response that enhances cell adhesion of *Escherichia coli*. Infect Immun. 2011 Oct;79(10):3895–904. Epub 2011 Jul 18.
2. Patel S, Mathivanan N, Goyal A. Bacterial adhesins, the pathogenic weapons to trick host defense arsenal. Biomed Pharmacother. 2017 Sep;93:763–771. Epub 2017 Jul 12.
3. Ichiro Nakagawa, Hiroaki Inaba, Taihei Yamamura. Invasion of epithelial cells and proteolysis of cellular focal adhesion components by distinct types of Porphyromonas gingivalis fimbriae.Infect Immun. 2006 Jul;74(7):3773–82.
4. Yoshida S, Inaba H, Nomura R, Murakami M et al. Efficacy of FimA antibody and clindamycin in silkworm larvae stimulated with Porphyromonas gulae. J Oral Microbiol. 2021 Apr 25;13(1):1914499.
5. Cecilia Ambrosi, Daniela Scribano. Acinetobacter baumannii Targets Human Carcinoembryonic Antigen-Related Cell Adhesion Molecules (CEACAMs) for Invasion of Pneumocytes. mSystems. 2020 Dec 22;5(6):e00604-20.
6. Cho SH, Lee KM, Kim CH, Kim SS. Construction of a Lectin-Glycan Interaction Network from Enterohemorrhagic *Escherichia coli* Strains by Multi-omics Analysis. Int J Mol Sci 2020, 21(8).
7. Seung-Hak Cho,Jun-young Park andCheorl-Ho Kim. Systemic Lectin-Glycan Interaction of Pathogenic Enteric Bacteria in the Gastrointestinal Tract. Int. J. Mol. Sci. 2022, 23(3), 1451; 27 Jan 2022.
8. Tenaillon O, Skurnik D, Picard B, Denamur E: The population genetics of commensal *Escherichia coli*. Nat Rev Microbiol 2010, 8(3):207–217.
9. Nataro JP, Kaper JB: Diarrheagenic *Escherichia coli*. Clin Microbiol Rev 1998, 11(1):142–201.
10. Caprioli A, Morabito S, Brugere H, Oswald E: Enterohaemorrhagic *Escherichia coli*: emerging issues on virulence and modes of transmission. Vet Res 2005, 36(3):289–311.
11. Riley LW, Remis RS, Helgerson SD, McGee HB, Wells JG, Davis BR, Hebert RJ, Olcott ES, Johnson LM, Hargrett NT et al: Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. N Engl J Med 1983, 308(12):681–685.
12. Beutin L: Emerging enterohaemorrhagic *Escherichia coli*, causes and effects of the rise of a human pathogen. J Vet Med B Infect Dis Vet Public Health 2006, 53(7):299–305.
13. O'Brien AD, Newland JW, Miller SF, Holmes RK, Smith HW, Formal SB: Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. Science 1984, 226(4675):694–696.
14. Karmali MA: Infection by Shiga toxin-producing *Escherichia coli*: an overview. Mol Biotechnol 2004, 26(2):117–122.
15. Sandvig K, Bergan J, Dyve AB, Skotland T, Torgersen ML: Endocytosis and retrograde transport of Shiga toxin. Toxicon 2010, 56(7):1181–1185.
16. S Langermann, S Palaszynski, M Barnhart. Prevention of mucosal *Escherichia coli* infection by FimH-adhesin-based systemic vaccination. Science. 1997 Apr 25;276(5312):607–11.
17. Perna NT, Plunkett G, 3rd, Burland V, Mau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gregor J, Kirkpatrick HA et al: Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature 2001, 409(6819):529–533.
18. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 2014 Jan;42(Database issue):D206-14. Epub 2013 Nov 29.
19. A Jafari, MM Aslani and S Bouzari. Escherichia coli: a brief review of diarrheagenic pathotypes and their role in diarrheal diseases in Iran. Iran J Microbiol. 2012 Sep; 4(3): 102–117
20. Denis Piérard, Henri De Greve, Freddy Haesebrouck, Jacques Mainil. O157:H7 and O104:H4 Vero/Shiga toxin-producing *Escherichia coli* outbreaks: respective role of cattle and humans. Vet Res. 2012 Feb 13;43(1):13.
21. Virag Sharma, Nikolai Hecker, Juliana G. Roscito, Leo Foerster, Bjoern E. Langer & Michael Hiller. A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. Nat Commun. 2018 Mar 23;9(1):1215.

22. Karam Mr Asadi, M Oloomi, M Habibi, S Bouzari. Cloning of *fimH* and *fliC* and expression of the fusion protein FimH/FliC from Uropathogenic *Escherichia coli* (UPEC) isolated in Iran. Iran J Microbiol. 2012 Jun;4(2):55–62.
23. Drew J Schwartz, Vasilios Kalas, Jerome S Pinkner et al. Positively selected FimH residues enhance virulence during urinary tract infection by altering FimH conformation. Proc Natl Acad Sci U S A. 2013 Sep 24;110(39):15530–7. Epub 2013 Sep 3.
24. Aprikian P, Tchesnokova V, Kidd B et al. Interdomain Interaction in the FimH Adhesin of *Escherichia coli* Regulates the Affinity to Mannose. J Biol Chem. 2007 Aug 10;282(32):23437–46. Epub 2007 Jun 13.
25. Meike Scharenberg, Xiaohua Jiang, Lijuan Pang. Kinetic properties of carbohydrate-lectin interactions: FimH antagonists. ChemMedChem. 2014 Jan;9(1):78–83. Epub 2013 Dec 2.
26. Laurel K Mydock-McGrane, Zachary T Cusumano, James W Janetka. Mannose-derived FimH antagonists: a promising anti-virulence therapeutic strategy for urinary tract infections and Crohn's disease. Expert Opin Ther Pat. 2016;26(2):175–97. Epub 2016 Jan 22.
27. Lee KM, Kim CH, Kim JH, Kim SS, Cho SH: e-Membranome: A Database for Genome-Wide Analysis of *Escherichia coli* Outer Membrane Proteins. Curr Pharm Biotechnol 2021, 22(4):501–507.
28. S Danner, J G Belasco. T7 phage display: a novel genetic selection system for cloning RNA-binding proteins from cDNA libraries. Proc Natl Acad Sci U S A. 2001 Nov 6;98(23):12954-9. Epub 2001 Oct 23.
29. Pablo San Segundo-Acosta, María Garranzo-Asensio, Carmen Oeo-Santos.High-throughput screening of T7 phage display and protein microarrays as a methodological approach for the identification of IgE-reactive components. J Immunol Methods. 2018 May;456:44–53. Epub 2018 Feb 20.
30. M Yamamoto, Y Kominato, F Yamamoto. Phage display cDNA cloning of protein with carbohydrate affinity. Biochem Biophys Res Commun. 1999 Feb 16;255(2):194–9.
31. Ielasi FS, Alioscha-Perez M, Donohue D, Claes S, Sahli H, Schols D, Willaert RG. Lectin-Glycan Interaction Network-Based Identification of Host Receptors of Microbial Pathogenic Adhesins. mBio. 2016 Jul 12;7(4):e00584-16.
32. Mariethoz J, Khatib K, Alocci D, Campbell MP, Karlsson NG, Packer NH, Mullen EH, Lisacek F. SugarBindDB, a resource of glycan-mediated host–pathogen interactions. Nucleic Acids Res. 2016 Jan 4;44(D1):D1243-50. Epub 2015 Nov 17.
33. Guinee PA, Agterberg CM, Jansen WH: *Escherichia coli* O antigen typing by means of a mechanized microtechnique. Appl Microbiol 1972, 24(1):127-131.7
34. Kwon T, Bak YS, Jung YH, Yu YB, Choi JT, Kim CH, Kim JB, Kim W, Cho SH. Whole-genome sequencing and comparative genomic analysis of *Escherichia coli* O91 strains isolated from symptomatic and asymptomatic human carriers. Gut Pathog. 2016 Nov 11;8:57. eCollection 2016.
35. Kwon T, Kim JB, Bak YS, Yu YB, Kwon KS, Kim W, Cho SH. Draft genome sequence of non-shiga toxin-producing *Escherichia coli* O157 NCCP15738. Gut Pathog. 2016 Apr 19;8:13. eCollection 2016.
36. Kwon T, Cho SH. Draft Genome Sequence of Enterohemorrhagic *Escherichia coli* O157 NCCP15739, Isolated in the Republic of Korea. Genome Announc. 2015 May 28;3(3):e00522-15.
37. Hackl T, Hedrich R, Schultz J, Forster F: proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. Bioinformatics 2014, 30(21):3004–3011.
38. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M et al: The RAST Server: rapid annotations using subsystems technology. BMC Genomics 2008, 9:75.
39. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q: VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res 2005, 33(Database issue):D325-328.
40. Khan NH, Ahsan M, Yoshizawa S, Hosoya S, Yokota A, Kogure K: Multilocus sequence typing and phylogenetic analyses of *Pseudomonas aeruginosa* Isolates from the ocean. Appl Environ Microbiol 2008, 74(20):6194–6205.
41. Glaeser SP, Kampfer P: Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. Syst Appl Microbiol 2015, 38(4):237–245.
42. Luz H Patiño, Milena Camargo, Marina Muñoz, Dora I Ríos-Chaparro, Manuel A Patarroyo, Juan D Ramírez.Unveiling the Multilocus Sequence Typing (MLST) Schemes and Core Genome Phylogenies for Genotyping *Chlamydia trachomatis*. Front Microbiol. 2018 Aug 22;9:1854. ECollection 2018.
43. Cho SH, Lee KM, Kim CH. Method for predicting outer membrane proteins of *escherichia coli* system thereof. Korea Patent 2021, No.1022429720000.

Tables

Table 1 to 3 are available in the Supplementary Files section.

Figures

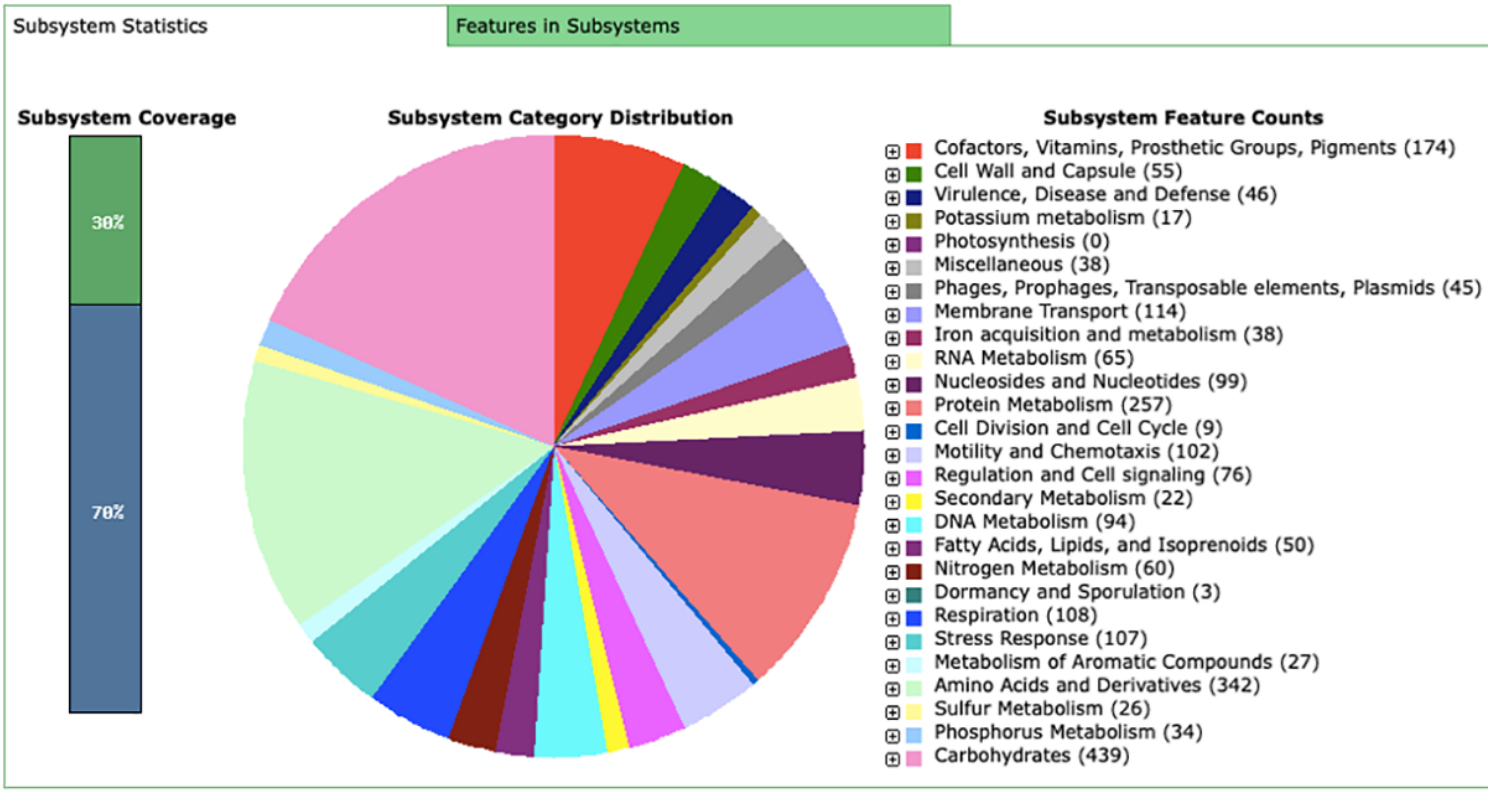

Figure 1

Subsystem category distribution of NCCP14539 based on SEED databases.

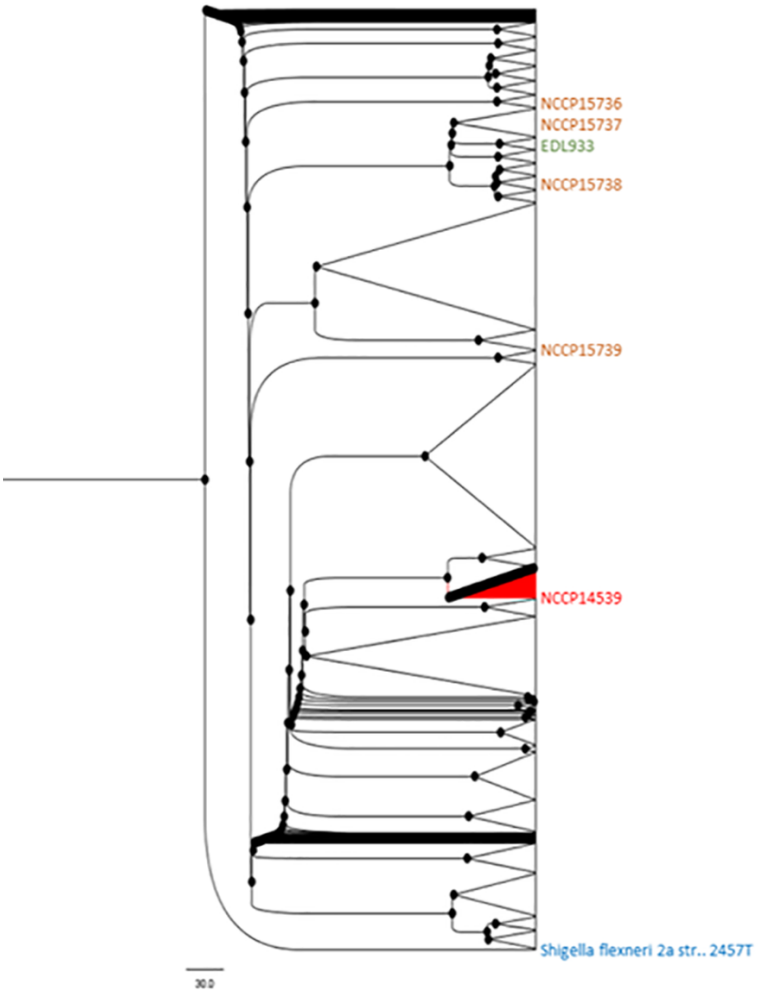

Figure 2

Multi-locus sequence typing (MLST)-based phylogenetic tree of NCCP14539. MLST-based phylogeny. Evolutionary time scaled by 100; lower values imply relatively recent branching. The scale indicates the number of substitutions per site. NCCP14539 did not belong to the E. coli O157:H7 serotype and was evolutionarily distant from E. coli O157:H7 str. EDL933 (green).

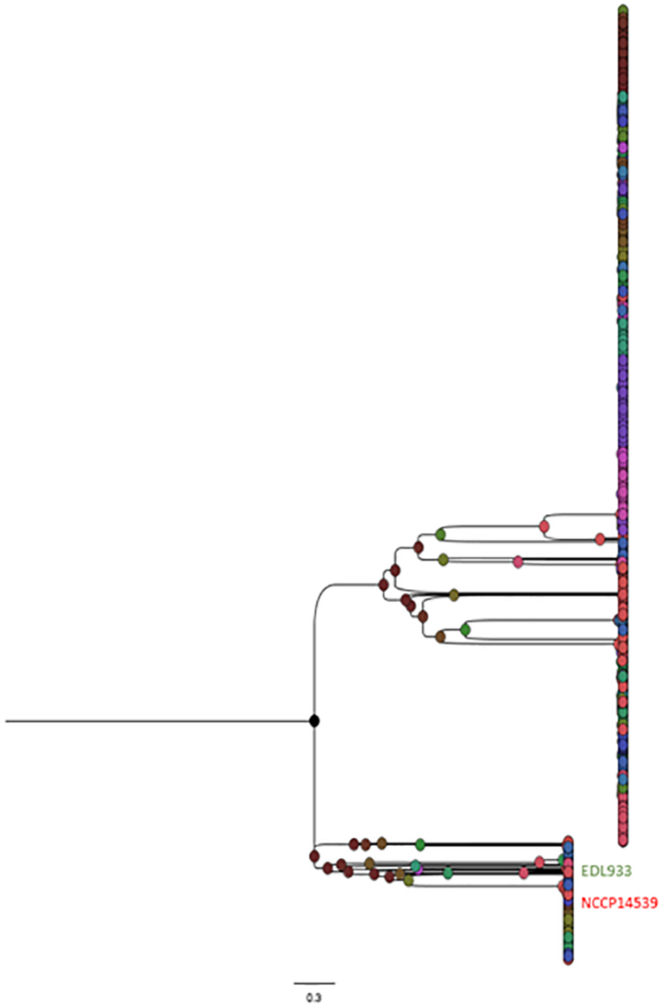

Figure 3

Phylogenetic tree of NCCP14539 and EHEC strains with respect to the fimH gene. Evolutionary time scaled by 100; lower values imply relatively recent branching. The scale indicates the number of substitutions per site. NCCP14539 belongs to the large phylogenetic group (red).

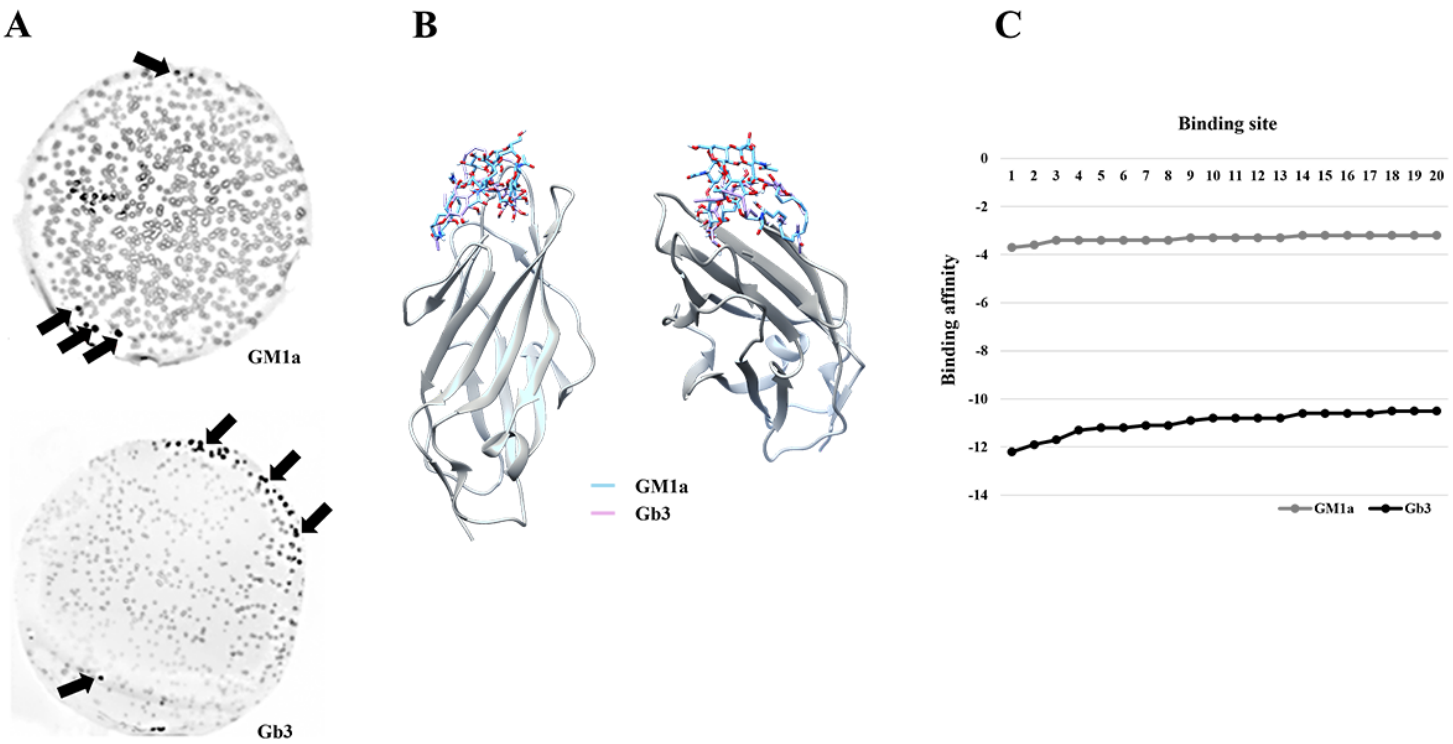

Figure 4

Phylogenetic tree of NCCP14539 and EHEC strains with respect to the fimH gene. Evolutionary time scaled by 100; lower values imply relatively recent branching. The scale indicates the number of substitutions per site. NCCP14539 belongs to the large phylogenetic group (red).

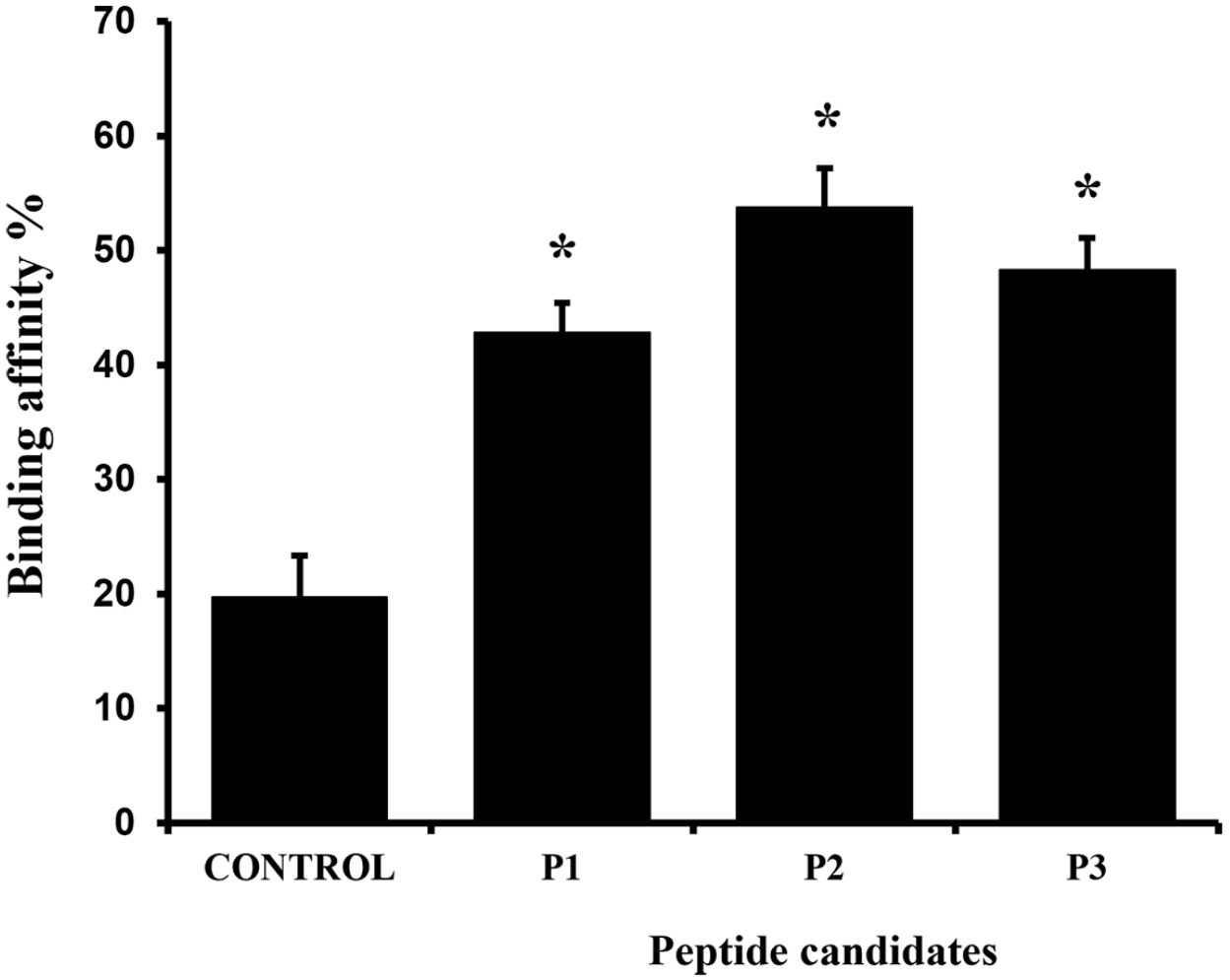

Figure 5

Elucidation of inhibitor restraining linkage of Pathogenic Enteric Bacteria EHEC NCCP14539 strain. A comparison of the binding affinities of Gb3-replica peptides to FimH in EHEC NCCP14539 was made. The term "control" refers to the binding affinity of a matrix without peptides. When compared to the Control, the P1–P3 peptides revealed statistically significant changes. *Means were substantially different (**P 0.05).

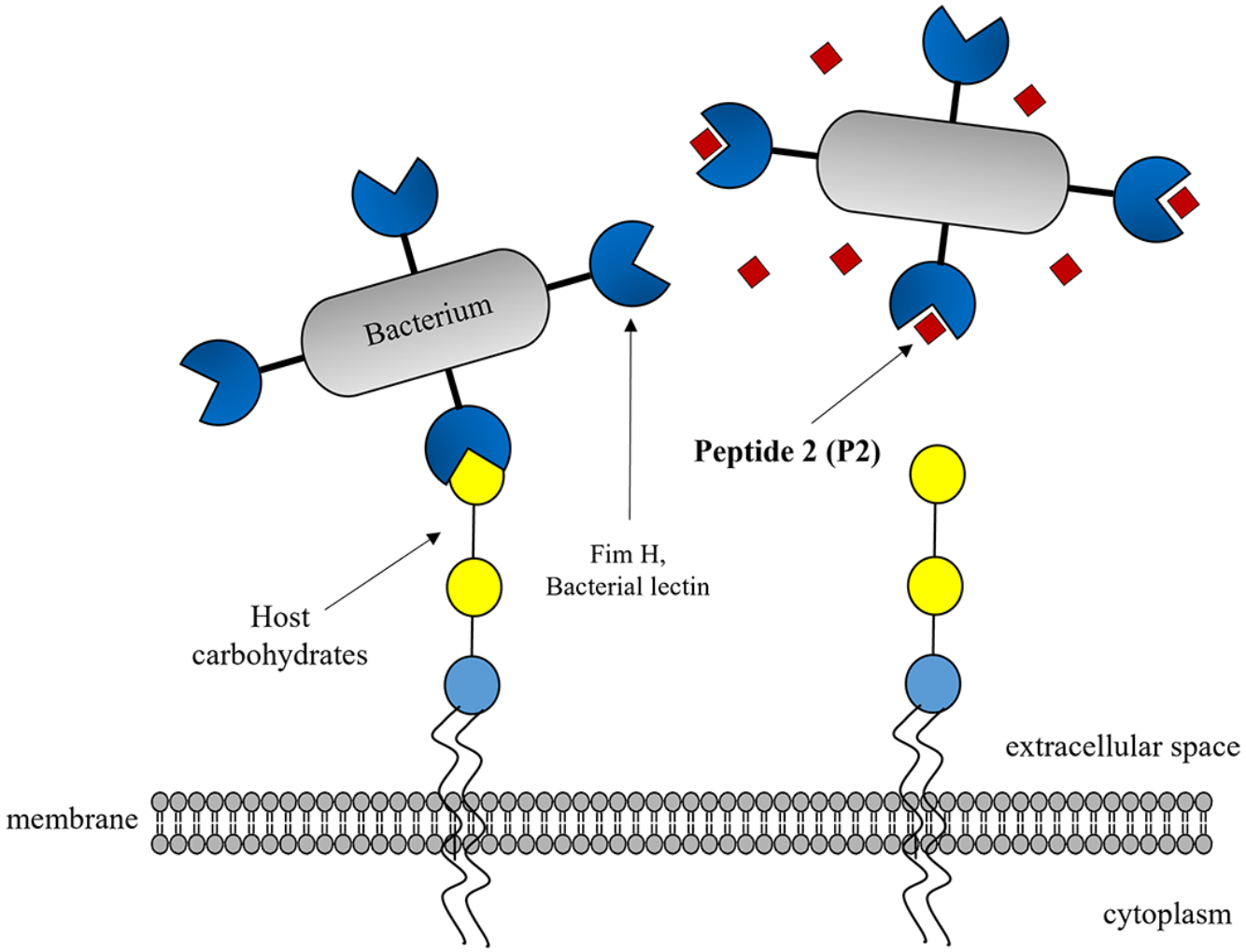

Figure 6

Diagrammatic illustration of elucidating an inhibitor-restraining connection in the EHEC NCCP14539 strain of pathogenic enteric bacteria. Immunotherapy for EHEC NCCP14539 strain might be used to prevent and cure infectious disorders by determining the involvement of the host Gb3 glycan peptide in the immune response against the pathogen.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.xlsx
- TableS2.xlsx
- Table.1.xlsx
- Table.2.xlsx
- Table.3.xlsx