Identification of a gene cluster for D-tagatose utilization in Escherichia coli B2 phylogroup

Jinyoung Ha, Dohyeon Kim, Jinho Yeom, Youngshin Kim, Seung Min Yoo, Sung Ho Yoon

syoon@konkuk.ac.kr

Highlights

- A gene cluster for D-tagatose utilization was found in Escherichia coli
- D-Tagatose catabolic pathway was experimentally identified
- Expression of the gene cluster changed the phenotype from Tag- to Tag+
- The gene cluster is prevalent in virulence-associated E. coli B2 phylogroup

Ha et al., iScience 25, 105655
December 22, 2022 © 2022
The Author(s)
https://doi.org/10.1016/j.isci.2022.105655

OPEN ACCESS
Identification of a gene cluster for D-tagatose utilization in *Escherichia coli* B2 phylogroup

Jinyoung Ha,1,3 Dohyeon Kim,1,3 Jinho Yeom,2 Youngshin Kim,1 Seung Min Yoo,2 and Sung Ho Yoon1,4,*

**SUMMARY**

D-Tagatose is a promising low-calorie sugar-substituting sweetener in the food industry. Most ingested D-tagatose is fermented by intestinal microorganisms. Until now, *Escherichia coli* has been considered incapable of growing on D-tagatose. Here, we discovered a gene cluster involved in D-tagatose utilization in *E. coli*. The chromosome of the intestinal probiotic *E. coli* Nissle 1917 contains a six-gene cluster encoding the ABC transporter, D-tagatose kinase, D-tagatose-bisphosphate aldolase, and putative aldose 1-epimerase. The functionality of the gene cluster was experimentally validated. Based on single-gene deletions, D-tagatose dissimilation occurs via D-tagatose 6-phosphate to D-tagatose 1,6-bisphosphate to D-glyceraldehyde 3-phosphate plus dihydroxyacetone phosphate. Remarkably, this gene cluster was located in 93% of the completely sequenced genomes of the *E. coli* B2 phylogroup, which contains the majority of extraintestinal pathogenic and adherent-invasive *E. coli* strains prevalent in patients with inflammatory bowel disease. This highlights the importance of understanding the clinical significance of D-tagatose in microbiota alterations.

**INTRODUCTION**

D-Tagatose, commonly called tagatose, is a C-4 epimer of D-fructose. It is a rare sugar naturally present in small amounts in dairy products and fruits1 and is generally recognized as safe (GRAS) by the US Food and Drug Administration (FDA).2 D-Tagatose is low in calories but tastes similar to sucrose. Owing to its bulk properties, D-tagatose has been widely used as a sweetener in a variety of foods and beverages. Various health and medical benefits have been discovered, including a low glycemic index, anti-plaque and anti-obesity properties, and reduction in symptoms associated with type 2 diabetes, hyperglycemia, anemia, and hemophilia.1 In particular, D-tagatose is receiving great attention for its potential to help manage type 2 diabetes and obesity.

Orally ingested D-tagatose is poorly digested by the host, and most (75–80%) are predominantly catabolized by intestinal bacteria.3 D-Tagatose treatment has a growth promoting effect on beneficial gut bacteria that produce short-chain fatty acids contributing to make the intestinal environment more harmonious.3–5 In the large intestine, only a few enteric microorganisms are known to be able to catabolize D-tagatose, most of which are lactic acid bacteria such as *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Enterococcus*, *Pediococcus*, and *Streptococcus*.6 In this regard, ingestion of D-tagatose appears to favorably modulate the composition of gut microbiota.3,5

Depending on the strain, D-tagatose can be either the sole carbon source or phosphorylated intermediates in the catabolism of lactose, D-galactose, and galactitol (Figure 1).7 In gram-positive bacteria such as *Staphylococcus aureus*8 and certain lactic acid bacteria,9 D-tagatose 6-phosphate (Tag 6-P) is generated as an intermediate metabolite in the Tag 6-P pathway (encoded by *lacABCD*) for the catabolism of lactose or D-galactose taken up by the sugar-specific phosphotransferase system (PTS). Similarly, in gram-negative Enterobacteriaceae species such as *E. coli*, *Klebsiella oxytoca*, *Klebsiella pneumoniae*, and *Salmonella enterica*, Tag 6-P is formed as an intermediate in the PTS-dependent galactitol catabolic pathway (encoded by *gatYABC*).10

The genomes of *K. oxytoca*, *K. pneumoniae*, and *S. enterica* contain an additional genes *tagKTH*, which is located between *gatY* and *gatZ* in the corresponding *gat* operon of *E. coli* (Figure 1). The tagKTH, together with *gatYZ*, encodes the PTS-mediated D-tagatose catabolic pathway.10 Extracellular D-tagatose is
transported into the cell by D-tagatose-PTS (TagTH) as D-tagatose 1-phosphate, which is converted to D-tagatose 1,6-bisphosphate (Tag 1,6-BP) by tagatose 1-phosphate kinase (TagK), followed by generation of D-glyceraldehyde 3-phosphate (G3P) plus dihydroxyacetone phosphate (DHAP) by heterodimeric tagatose 1,6-bisphosphate (TBP)-specific aldolase (encoded by GatYZ). A pathway closely related to the PTS-mediated D-tagatose catabolic pathway has been reported for the gram-positive bacterium *Bacillus licheniformis*. A gene cluster for the catabolism of D-tagatose and galactitol using an ATP-binding cassette (ABC) transporter, instead of the PTS transporter, was found on the symbiotic plasmid in the gram-negative legume endosymbiont *Sinorhizobium meliloti* (Figure 1). The gene cluster encodes the ABC-mediated D-tagatose catabolic pathway. D-Tagatose is taken up by ABC transfer (TagABC), which is phosphorylated to generate Tag 6-P by tagatose kinase (TagD), followed by generation of D-fructose-6-phosphate by Tag 6-P epimerase (TagE).

Gut microbiota is known to play a key role in human health by modulating metabolic capabilities, immune responses, and resistance to pathogenic infections. Inflammatory bowel disease (IBD) is thought to be...
associated with compositional alteration of the intestinal microbiota with reduced diversity in major phyla such as Firmicutes and Bacteroidetes and increased numbers of Enterobacteriaceae such as E. coli phy- 
groups B2 and D.\textsuperscript{13,14} The gene repertoires of B2 strains were found to be the most distinct from those of the other major phylogroups A, B1, and D and the most diverse among the phylogroups.\textsuperscript{13,14} B2 strains exhibit a variety of disease states, ranging from pathogenic to probiotic, and are differently associated with certain ecological niches in the human body and propensity to cause disease. The B2 phylogroup includes most extraintestinal pathogenic E. coli (ExPEC) strains\textsuperscript{12,13} as well as most adherent-invasive E. coli (AIEC) strains that have been proposed as triggers of IBD.\textsuperscript{12} The probiotic E. coli strain Nissle 1917 (EcN), a B2 strain, has strong antagonistic activity against entero-pathogens, possesses immunomodulatory properties, and has been widely used for IBD treatment.\textsuperscript{20,21} The metabolic potential of EcN is similar to that of uropathogenic B2 strain CFT073.\textsuperscript{22,23}

Until now, E. coli has been known to be unable to utilize D-tagatose as the sole carbon source, due to the lack of tagKTH.\textsuperscript{10} However, in a previous comparative phenome analysis,\textsuperscript{23} we observed that EcN grew on D-tagatose as the sole carbon source, whereas intestinal commensal E. coli K-12 did not. In the present study, we identified a novel gene cluster required for D-tagatose utilization in EcN. Supplementation experiments were performed to determine the functionality of the gene cluster. The metabolic pathway associated with the gene cluster has also been experimentally identified. Genomic regions homologous to the gene cluster were searched throughout all the bacterial sequenced genomes, which indicated the remarkable prevalence of the D-tagatose gene cluster in E. coli strains of the B2 phylogroup.

RESULTS

Identification of a novel gene cluster for D-tagatose utilization in EcN
In a previous comparative phenome analysis,\textsuperscript{23} we found that E. coli K-12 MG1655 did not grow using D-tagatose as the sole carbon source, whereas the probiotic EcN did. Thus, we compared the genome sequences of EcN (RefSeq: NZ_CP007799)\textsuperscript{24} and K-12 MG1655 (RefSeq: NC_000913)\textsuperscript{25} to identify the EcN-specific genomic region involved in D-tagatose catabolism. The candidate region was completely absent in K-12 and was located between accC (ECOLIN_RS18665) encoding biotin carboxylase and yhdT (ECOLIN_RS18710) encoding a conserved inner membrane protein (Figure 2A).

The 8.5 kb region contains eight protein-coding genes (locus tags ECOLIN_RS18670 to ECOLIN_RS18705) that are organized into two seemingly operons separated by 341 bp. The former gene cluster contained six genes on the reverse strand that were annotated to encode tagatose bisphosphate family class II aldolase (encoded by ECOLIN_RS18695), substrate-binding domain-containing protein (ECOLIN_RS18690), sugar ABC transporter ATP-binding protein (ECOLIN_RS18685), ABC transporter permease (ECOLIN_RS18680), sugar kinase (ECOLIN_RS18675), and aldose 1-epimerase family protein (ECOLIN_RS18670). The latter gene cluster on the forward strand contained the remaining two genes, encoding the DeoR/GlpR transcriptional regulator (ECOLIN_RS18700) and sugar kinase (ECOLIN_RS18705). A plausible metabolic route encoded by the former gene cluster is shown in Figure 2B.

To investigate the effect of initial concentration of D-tagatose on cell growth, we grew EcN in M9 minimal medium supplemented with different concentrations (zero to 500 mM) of D-tagatose for 120 h (Figure S1). Up to 40 mM of D-tagatose, the final cell density increased without residual D-tagatose at the end of the culture. Afterward, the final cell density decreased with remaining of D-tagatose. In the case of 500 mM D-tagatose, the cells growth was severely repressed, and the added D-tagatose was hardly consumed. Based on this result, 40 mM of D-tagatose was used for cell cultures using D-tagatose as the sole carbon source.

Gene expression analysis
For wild-type EcN cells grown on D-tagatose, the two gene clusters were subjected to quantitative real-time PCR (qRT-PCR) analysis (Figure 2C). Compared to cells grown on glucose as the sole carbon source, in cells grown on D-tagatose, the relative expression levels of all six genes (ECOLIN_RS18670 to ECOLIN_RS18695) in the former gene cluster increased significantly (>40-fold, p value <0.01), especially for the first three genes (ECOLIN_RS18685, ECOLIN_RS18690, and ECOLIN_RS18695) (>1000-fold). Growth on D-tagatose also resulted in higher transcriptional levels of ECOLIN_RS18700 and ECOLIN_RS18705 in the latter gene cluster, but to a much lesser extent (around two-fold, p value <0.05). The markedly high expression upon D-tagatose reflects the former gene cluster is strictly regulated
in response to the availability of the uncommon nutrient. The different gene expression levels and directions of transcription indicated that the two gene clusters formed separate operons. Gene annotation and qRT-PCR analysis suggest that the former gene cluster plays an important role in D-tagatose utilization compared with the latter gene cluster.

**Growth of E. coli strains on D-tagatose by expression of EcN tagatose gene cluster**

The function of the former gene cluster (ECOLIN_RS18670 to ECOLIN_RS18695) in D-tagatose utilization was explored using complementation experiments (Figure 3A). Deletion of the gene cluster in the EcN chromosome (referred to as EcNΔ) resulted in no growth with D-tagatose (Tag−). Cell growth on D-tagatose (Tag+) was fully restored to a level comparable to that of wild-type EcN by complementation with the pHCE-6 construct, which constitutively expresses the former gene cluster (EcNΔ/pHCE-6). Two genes, located on the gene clusters of the former (ECOLIN_RS18675, 963 bp in length) and the latter (ECOLIN_RS18705, 936 bp), were annotated as sugar kinase, and they showed a high protein sequence identity of 51%. Thus, we were intrigued to find that the in-frame, non-polar deletion of

---

**Figure 2. Gene cluster for D-tagatose degradation in E. coli Nissle 1917 (EcN)**

(A) Genetic organization of the six-membered gene cluster on the reverse strand (ECOLIN_RS18670 to ECOLIN_RS18695 in different colors; referred to as EcN tag cluster) and two-membered gene cluster on the forward strand (ECOLIN_RS18700 and ECOLIN_RS18705). Arrows denote direction of transcription, and genes with similar functions are indicated by the same color. This genomic region is missing between accC and yhdT in the genome of E. coli K-12 MG1655. Picture was drawn to scale.

(B) Proposed D-tagatose utilization pathway associated with EcN tag cluster. Gene(s) corresponding to each metabolic step (arrow) are boxed: genes from EcN tag cluster are colored and genes from the rest of EcN genome are in italic.

(C) Relative gene expression levels in EcN growing on D-tagatose or glucose. Wild-type cells were cultured in defined medium with 3 g/L glucose or 40 mM D-tagatose, and were sampled at OD600 of 1.0. The expression levels of cells growing on glucose were set as a relative expression of 1. Data are presented as log10-transformed relative expression levels. Error bars represent the standard error of the mean (SEM) from three independent cultivations with three replicates. Asterisks denote significant difference from the growth on glucose (*p value < 0.05, **p value < 0.01). For clarity, the locus tags (ECOLIN_) were removed from the locus names.
ECOLIN_RS18675 in the EcN chromosome (EcN D1) led to the Tag+/C0 phenotype (Figure 3A). This indicates that ECOLIN_RS18675 in the former gene cluster should encode tagatose kinase that converts D-tagatose to Tag 6-P. The concentration of D-tagatose added to the culture medium decreased with increasing cell density in cultures of EcN and EcN D6 and EcN D1. The accumulation of by-products, such as acetic and formic acids, was negligible throughout the cultivation of all strains.

To corroborate the capability of the former gene cluster, we introduced pHCE-6 into various laboratory wild-type E. coli strains, K-12 (MG1655 and BW25113), B (BL21(DE3)), and W, which cannot grow on D-tagatose. Expression of the former gene cluster changed the phenotype from Tag+/C0 to Tag+ (Figures 3B and 4). In all cells with the Tag+ phenotype, the D-tagatose concentration in the culture media decreased with cell growth, and no D-tagatose was detected at the end of cultures. These results clearly demonstrate that the six-membered gene cluster is involved in the D-tagatose utilization in E. coli strains. Hereafter, the former gene cluster will be referred to as EcN tag cluster.

To investigate the potential of EcN tag cluster to utilize D-tagatose under anaerobic or hypoxic condition, we performed a complementation experiment under anaerobic condition using M9 medium supplemented with D-tagatose and yeast extract for 120 h (Figure 5). As controls, cells were cultured using M9 medium supplemented only with yeast extract. Wild-type EcN and EcN ΔECOLIN_RS18670 to ECOLIN_RS18695 did not grow on D-tagatose without consumption of the added D-tagatose. This result clearly shows that EcN tag cluster can utilize D-tagatose under anaerobic or hypoxic condition.

**Identification of the D-tagatose utilization pathway associated with EcN tag cluster**

Gene annotation of EcN tag cluster suggests the ABC-mediated D-tagatose degradation pathway (Figure 2B): extracellular D-tagatose → cytosolic D-tagatose → Tag 6-P → Tag 1,6-BP → G3P + DHAP. To complete this route, EcN tag cluster is missing a gene encoding Tag 6-P kinase that converts Tag 6-P to Tag 1,6-BP. This reaction is mainly catalyzed by 6-phosphofructokinase I (encoded by pfkA), as well as 6-phosphofructokinase II (pfkB) which is usually expressed at low levels in E. coli. The conversion of Tag 1,6-BP to G3P plus DHAP can be performed by TBP aldolase which is encoded in EcN tag cluster.
In addition, *E. coli* possesses *gatYZ* and *kbaYZ* encoding two distinct TBP aldolases (GatY and KbaY) involved in galactitol catabolism.\(^{29}\) GatZ and KbaZ have no aldolase activity, however, are required for full activity of GatY and KbaY, respectively. Global pairwise alignment of the protein sequences showed that ECOLIN_RS18695 (284 amino acids) was more similar to GatY (284 aa, 63.7% identity) than to KbaY (286 aa, 56.3%).

To validate this metabolic route, we expressed EcN tag cluster in the background of deletion of metabolic enzymes (PfkA or PfkB) proposed to catalyze the missing metabolic step (Figure 3B). To this end, we introduced pHCE-6 into in-frame deletion mutants (BW25113 \(\text{D}^{\text{pfkA}}\) and BW25113 \(\text{D}^{\text{pfkB}}\)) obtained from Keio collection of single-gene deletion mutants of *E. coli* K-12 BW25113.\(^{30}\) Gene deletions were confirmed by PCR. As negative controls, none of the deletion mutants or their parental strain (BW25113) grew on D-tagatose. As a positive control, BW25113 harboring pHCE-6 (BW25113/pHCE-6) grew on D-tagatose; however, BW25113 \(\text{D}^{\text{pfkA}}\)/pHCE-6 did not grow on D-tagatose without consuming any D-tagatose added to the medium. These results demonstrate that PfkA is essential for D-tagatose utilization, and that the EcN tag cluster, together with PfkA, comprises the metabolic route depicted in Figure 2B.

Phylogenetic distribution of EcN tag cluster in bacterial genomes

To analyze the distribution of genomic regions homologous to the eight-gene cluster of EcN within each of the 24,025 bacterial complete genomes, their proteomes were screened using BLASTP. The genomes analyzed covered a broad spectrum of bacterial diversity and had several different levels of intragroup divergence. We found that a total of 490 genomes contained a gene cluster containing at least four homologs of tagatose kinase (ECOLIN_RS18675) and ABC transport subunits (ECOLIN_RS18680, ECOLIN_RS18685, and ECOLIN_RS18690) (Figure 6). The cladogram describing the phylogenetic relationships among them was constructed based on the NCBI Taxonomy database and the phylogenetic tree of each phylum.\(^{31-44}\) The 490 genomes were classified into 55 bacterial species. Most belonged to the phylum Proteobacteria (477 genomes of 46 species), followed by Firmicutes (11 out of 7 species), Fusobacteria (1 ea), and Spirochaetes (1 ea).

The majority of the genomes of the phylum Proteobacteria belonged to class \(\gamma\)-Proteobacteria (403 genomes of 11 species), followed by \(\alpha\)-Proteobacteria (57 genomes of 23 species), and \(\beta\)-Proteobacteria (17 genomes of 12 species). Of the 403 genomes of class \(\gamma\)-Proteobacteria, the family Enterobacteriaceae accounted for 96.8% (390 genomes), whereas families Yersiniaceae (8 genomes), Vibrionaceae (2),
Bruguerivoracaceae (1), Oceanospi rillaceae (1), and Cardiobacteriaceae (1) combined constituted only 3.2%. The EcN genomic region was fully conserved in all matched genomes of the family Enterobacteri-aceae (390 genomes), most of which were E. coli strains (385). Of interest, ECOLIN_RS18700 and ECOLIN_RS18705 were found only in the order Enterobacterales, not in other members of the order, sug-gesting that these genes play a limited role in D-tagatose utilization.

Four homologs of tagatose kinase and ABC transport subunits were well conserved in the genomes of a-Proteobacteria (57 ea) and b-Proteobacteria (17), although ECOLIN_RS18670, encoding putative aldose 1-epimerase, was replaced by a Tag 6-P epimerase gene. In addition to these genes, homologs of the DeoR/GlpR transcriptional regulator gene (ECOLIN_RS18700) contained in the latter gene cluster were observed in these homologous clusters. Among the homologous clusters, homologs of the TBP aldolase gene (ECOLIN_RS18695) were found only in class g-Proteobacteria and phylum Firmicutes. Of interest, they were found in gram-positive Firmicutes of Clostridia and Erysipelotrichia isolates but not in gram-negative Negativicutes class, and their locations varied from species to species.

A total of 451 homologous regions were found on the chromosomes, and 39 were found on extra-chromo-somal elements such as plasmids (Table S1). Overall, the average absolute deviation of their G + C content compared to that of their own genome was 1.9%, with a high degree of variation (standard deviation of 1.78%). Notably, the homologous regions found in Halothermothrix orenii H 168 and Moorella thermoace-tica showed considerably higher (4.9%) and lower (–8.1%) G + C contents, respectively. The homologous regions found in members of the family Enterobacteriaceae had lower G + C contents (–4.9% on average) and were otherwise absent from the genomes of family Enterobacteriaceae at the accC/yhdT locus (Figure 2A). This may indicate that these gene clusters were acquired via horizontal gene transfer.

### Distribution of EcN tag cluster in E. coli phylogroups

Genomic regions homologous to EcN tag cluster (385 out of 490) were found predominantly in E. coli genomes (Figure 6), and we further determined whether they were over-represented in specific phylogroups of the E. coli strains. To this end, in silico E. coli phylotyping using ClermonTyping software was performed for all E. coli genomes (2017 ea) subjected to homology searches. The E. coli genomes analyzed were classified into 14 phylogroups: A (621 ea), B1 (435), B2 (318), C (63), D (185), E (294), F (62), G (25), E or clade I (2), clade I (2), clade II (1), clade III (2), none Escherichia (2), and unknown (5). The 385 homologous regions were found in seven phylogroups, with the majority (297; 93.4% frequency) belonging to phy-logroup B2, followed by F (45; 72.6%), B1 (29; 6.7%), D (6; 3.2%), A (3; 0.5%), E (2; 0.7%), clade I (2; 100%),
and unknown (1; 20%). Considering that the F phylogroup is closely related to the B2 phylogroup, 47,48 this enrichment analysis indicated that the EcN tag cluster was confined and significantly concentrated (hypergeometric p value: $1.4 \times 10^{-244}$) in B2 phylogroup.

To confirm over-representation of EcN tag cluster in the B2 phylogroup, we performed a homology search using BLASTN for (in)complete genomes of 110 E. coli strains listed in a previous study 49 (Figure 7 and Table S2). The 110 E. coli strains included 53 isolates from IBD patients and healthy subjects: 55 were pathogens and 55 were commensals, based on pathotype. Of these, 66 belonged to the B2 phylogroup, whereas 44 other strains represented phylogroups A (16 ea), B1 (10), D (3), E (5), F (1), or unknown (1), and Shigella (8). The entire EcN tag cluster was found in 66 E. coli strains. Remarkably, 63 of the 66 B2 strains (95%) possessed the whole gene cluster, among which the ECOLIN_RS18670 homolog in H305 and ECOLIN_RS18705 homologs in S88 and H252 were annotated as pseudogenes that appear to require confirmation. Only three B2 strains, O127:H6 E2348/69, 41-2-TI13, and 63-1-TI1, possessed no homologs of EcN tag cluster. Homologous regions were also found in phylogroups B1 (two ea) and D (one ea). Among the eight Shigella strains that in fact belong to the E. coli species, 50 two S. boydii strains contained parts of homologs of EcN tag cluster.

The highly biased phylogenetic distribution highlights the fact that the presence of the EcN tag cluster is a genetic feature of the B2 phylogroup.
In microbial catabolism, phosphorylated forms of D-tagatose are well-known intermediates of the tagatose-6-phosphate pathway, which is responsible for lactose and galactose degradation in certain gram-positive bacteria, including lactic acid bacteria (Figure 1). D-Tagatose can also be used as the sole carbon source in certain species of Enterobacteriaceae such as K. oxytoca and S. enterica. Until now, E. coli has been considered incapable of growing on D-tagatose. A previous comparative genome analysis of transport proteins reported that probiotic and pathogenic E. coli strains, including EcN, do not have any transport system for D-tagatose. In this study, we identified a gene cluster required for D-tagatose degradation in probiotic EcN and the related catabolic pathway (Figure 2). The EcN tag cluster was shown to be associated with a functional metabolic route leading to the EMP pathway: extracellular D-tagatose → cytosolic D-tagatose → Tag 6-P → Tag 1,6-BP → G3P + DHAP. Among the gene clusters encoding D-tagatose-associated pathways reported to date (Figure 1), the EcN tag cluster resembled that of the rhizobacterium S. meliloti containing the ABC transporter gene. To date, PTS-dependent pathways have been considered major routes for the catabolism of D-tagatose by enteric bacteria. To our knowledge, the ABC transporter for D-tagatose has not been found in enteric bacteria, although it has only been reported for rhizobacteria, plant growth-promoting rhizobacteria S. meliloti and the tumor-producing plant pathogen Agrobacterium tumefaciens C58. D-Tagatose gene clusters of EcN and S. meliloti shared four genes encoding ABC transporter and tagatose kinase, and had structural similarities with each other. Nevertheless, the pathways involved differ in that the phosphorylated intermediate (Tag 6-P) is converted to Tag 1,6-BP in EcN, whereas it is converted to D-fructose 6-phosphate in S. meliloti. Intriguingly, EcN tag cluster contained ECOLIN_RS18670 encoding putative aldose 1-epimerase at the corresponding position of tagE encoding Tag 6-P epimerase in the S. meliloti tag cluster, although the sequences of ECOLIN_RS18670 and tagE were highly divergent. The functionality of ECOLIN_RS18670 warrants further investigation. Notably, EcN tag cluster contained the TBP aldolase gene (ECOLIN_RS18695), in addition to the known TBP aldolase genes (gatY and kbaY) located in the rest of the EcN genome. TBP aldolase is involved in utilizing sugars derived from intestinal mucus, and the presence of additional TBP aldolases has been suggested to be beneficial for intestinal colonization.

In this study, we identified a gene cluster required for D-tagatose degradation in probiotic EcN and the related catabolic pathway (Figure 2). The EcN tag cluster was shown to be associated with a functional metabolic route leading to the EMP pathway: extracellular D-tagatose → cytosolic D-tagatose → Tag 6-P → Tag 1,6-BP → G3P + DHAP. Among the gene clusters encoding D-tagatose-associated pathways reported to date (Figure 1), the EcN tag cluster resembled that of the rhizobacterium S. meliloti containing the ABC transporter gene. To date, PTS-dependent pathways have been considered major routes for the catabolism of D-tagatose by enteric bacteria. To our knowledge, the ABC transporter for D-tagatose has not been found in enteric bacteria, although it has only been reported for rhizobacteria, plant growth-promoting rhizobacteria S. meliloti and the tumor-producing plant pathogen Agrobacterium tumefaciens C58. D-Tagatose gene clusters of EcN and S. meliloti shared four genes encoding ABC transporter and tagatose kinase, and had structural similarities with each other. Nevertheless, the pathways involved differ in that the phosphorylated intermediate (Tag 6-P) is converted to Tag 1,6-BP in EcN, whereas it is converted to D-fructose 6-phosphate in S. meliloti. Intriguingly, EcN tag cluster contained ECOLIN_RS18670 encoding putative aldose 1-epimerase at the corresponding position of tagE encoding Tag 6-P epimerase in the S. meliloti tag cluster, although the sequences of ECOLIN_RS18670 and tagE were highly divergent. The functionality of ECOLIN_RS18670 warrants further investigation. Notably, EcN tag cluster contained the TBP aldolase gene (ECOLIN_RS18695), in addition to the known TBP aldolase genes (gatY and kbaY) located in the rest of the EcN genome. TBP aldolase is involved in utilizing sugars derived from intestinal mucus, and the presence of additional TBP aldolases has been suggested to be beneficial for intestinal colonization.

The phylogenetic analysis revealed that the majority of EcN tag cluster homologs were found in the class γ-Proteobacteria. In particular, all genes in EcN tag cluster were well-conserved in strains belonging to the γ-Proteobacteria.
family Enterobacteriaceae, accounting for 96.8% of the homologous gene clusters found in the bacterial genomes (390/403) (Figure 6). Their nucleotide composition differed from that of the rest of their genomes (~4.9% of A + G + C on average). Analysis of various E. coli phylogroups showed that 95% of the B2 strains contained EcN tag clusters between the accC and yhdT loci, which tended to be absent in non-B2 strains (Figure 7). The G + C content anomaly and lineage-specific abundance of EcN tag clusters indicates early genetic acquisition via recent horizontal gene transfer.

The gut microbiota is profoundly affected by what we eat,53 and certain food additives can be detrimental to human health. For example, consumption of artificial sweeteners was reported to cause functional and compositional changes in the gut microbiota, leading to induce glucose intolerance.54 D-Tagatose has been widely and unrestrictedly used as a food additive, and is considered a promising prebiotic with growth-promoting effects on beneficial gut bacteria.5,3 However, the prevalence of the EcN tag cluster in the B2 phylogroup (Figure 6) suggests that D-tagatose utilization might contribute to the inherent fitness of B2 strains, potentially leading to the overgrowth of the virulence-associated B2 phylogroup over other phylogroups in the gut microbiota. This implies that the administration of D-tagatose requires special caution in individuals suffering from chronic intestinal inflammation, such as IBD, a disease that is growing exponentially in westernized countries. This safety concern of critical importance should be investigated through microbiome studies and clinical trials, given the great potential of D-tagatose in the food and pharmaceutical industries.

Limitations of the study
In this study, a single deletion of tagatose kinase gene (ECOLIN_RS18675) in EcN tag cluster was made and tested. However, we have not made and characterized mutations in each of the genes in the gene cluster individually. For example, the possible role of the transcriptional regulator (ECOLIN_RS18700) in mediating up-regulation of the genes by growth on tagatose has not been tested, nor has the function of the putative aldose epimerase (ECOLIN_RS18670). The complete genetic analysis of the gene cluster and biochemical verification by way of activity assays could be potential directions for future study.

STAR METHODS
Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Bacterial strains
- METHOD DETAILS
  - Cell culture
  - Analytical procedures
  - Construction of gene deletion mutants
  - Construction of an expression vector
  - Quantitative real-time PCR (qRT-PCR)
  - Homology search
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION
Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.105655.

ACKNOWLEDGMENTS
This paper was supported by Konkuk University in 2018.

AUTHOR CONTRIBUTIONS
S.H.Y. conceived, organized, and supervised the project. J.H. performed the cell cultivation and genetic manipulation. D.K. performed bioinformatics analyses. Y.K. contributed to the genetic manipulations.
We support inclusive, diverse, and equitable conduct of research.

The authors declare no competing interests.

S.H.Y. wrote the manuscript. All the authors have read and approved the final manuscript.

REFERENCES

1. Oh, D.K. (2007). Tagatose: properties, applications, and microbiological processes. Appl. Microbiol. Biotechnol. 76, 1–8. https://doi.org/10.1007/s00253-007-0981-1.

2. Levin, G.V. (2002). Tagatose, the new GRAS sweetener and health product. J. Med. Food 5, 23–36. https://doi.org/10.1089/109662002753723197.

3. Roy, S., Chikkerur, J., Roy, S.C., Dhali, A., Kolte, A.P., Sridhar, M., and Samanta, A.K. (2018). Tagatose as a potential nutraceutical: production, properties, biological roles, and applications. J. Food Sci. 83, 2699–2709. https://doi.org/10.1111/1750-3841.14358.

4. Laerke, H.N., Jensen, B.B., and Højsgaard, S. (2000). In vitro fermentation pattern of D-tagatose is affected by adaptation of the microbiota from the gastrointestinal tract of pigs. J. Nutr. 130, 1772–1779. https://doi.org/10.1093/jn/130.7.1772.

5. Liang, Y.X., Wen, P., Wang, Y., OuYang, D.M., Wang, D., Chen, Y.Z., Song, Y., Deng, J., Sun, Y.M., and Wang, H. (2019). The constipation-relieving property of D-tagatose by modulating the composition of gut microbiota. Int. J. Mol. Sci. 20, 5721. https://doi.org/10.3390/ijms2025721.

6. Hans Bertelsen, H.A.M.T. (2001). Fermentation of D-tagatose by human intestinal bacteria and dairy lactic acid bacteria. Microb. Ecol. Health Dis. 13, 87–95. https://doi.org/10.1080/08910600119905.

7. Van der Heiden, E., Delmarcelle, M., Lebrun, S., Freichels, R., Brans, A., Vastenavond, C.M., Galleni, M., and Joris, B. (2013). A pathway closely related to the D-tagatose pathway of gram-negative enterobacteria identified in the gram-positive bacterium Bacillus licheniformis. Appl. Environ. Microbiol. 79, 3511–3515. https://doi.org/10.1128/aem.03918-12.

8. Rosey, E.L., Oskouian, B., and Stewart, G.C. (1991). Lactose metabolism by Staphylococcus aureus: characterization of lacABC, the structural genes of the tagatose 6-phosphate pathway. J. Bacteriol. 173, 5992–5998. https://doi.org/10.1128/jb.173.19.5992-5998.1991.

9. Kandler, O. (1983). Carbohydrate metabolism in lactic acid bacteria. Antonie Leeuwenhoek 49, 209–224. https://doi.org/10.1007/bf00399499.

10. Shakeri-Garakani, A., Brinkkötter, A., Schmid, K., Turtug, S., and Lengeler, J.W. (2004). The genes and enzymes for the catabolism of galactitol, D-tagatose, and related carbohydrates in Klebsiella oxytoca M5A1 and other enteric bacteria display convergent evolution. Mol. Genet. Genom. 271, 717–728. https://doi.org/10.1007/s00438-004-1022-8.

11. Kohlmeier, M.G., White, C.E., Fowler, J.E., Finan, T.M., and Orennik, I.J. (2019). Galactitol catabolism in Sinorhizobium meliloti is dependent on a chromosomally encoded sorbitol dehydrogenase and a pSymB-encoded operon necessary for tagatose catabolism. Mol. Genet. Genom. 294, 739–755. https://doi.org/10.1007/s00438-019-01545-z.

12. Pickard, J.M., Zeng, M.Y., Caruso, R., and Kolte, A.P., Sridhar, M., and Samanta, A.K. (2005). Genomic insights into an old probiotic bacterium. Dig. Bowel Dis. 15, 70–89. https://doi.org/10.1111/imr.12567.

13. Maloy, K.J., and Powrie, F. (2011). Intestinal microbiota: role in pathogen colonization, immune responses, and inflammatory disease. Immunol. Rev. 244, 739–755. https://doi.org/10.1111/j.0904-9745.2010.00538.x.

14. Kotlowski, R., Bernstein, C.N., Sepehri, S., and Krause, D.O. (2007). High prevalence of Escherichia coli belonging to the B2-D phylogenetic group in inflammatory bowel disease. Gut 56, 669–675. https://doi.org/10.1136/gut.2006.099796.

15. Touchon, M., Perrin, A., de Sousa, J.A.M., Vangchhia, B., Burn, S., O’Brien, C.L., Denamur, E., Gordon, D., and Rocha, E.P. (2008). Phylogenetic background and habitat drive the genetic diversification of Escherichia coli. PLoS Genet. 4, e1000374. https://doi.org/10.1371/journal.pgen.1000374.

16. Touchon, M., Hœder, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., et al. (2009). Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. PLoS Genet. 5, e1000344. https://doi.org/10.1371/journal.pgen.1000344.

17. Moulin-Schouleur, M., Répérant, M., Laurent, S., Brée, A., Mignon-Graustein, S., Germon, P., Rasschaert, D., and Schouler, C. (2007). Extraintestinal pathogenic Escherichia coli strains of avian and human origin: link between phylogenetic relationships and common virulence patterns. J. Clin. Microbiol. 45, 3366–3376. https://doi.org/10.1128/jcm.00370-07.

18. Picard, B., Garcia, J.S., Gouriou, S., Duriez, P., Brahim, N., Bingen, E., Elion, J., and Denamur, E. (1999). The link between phylogeny and virulence in Escherichia coli extraintestinal infection. Infect. Immun. 67, 546–553. https://doi.org/10.1128/iai.67.2.546-553.1999.

19. Martinez-Medina, M., Aldeguer, X., Lopez-Siles, M., Gonzalez-Huix, F., Lopez-Oliu, C., Dahbi, G., Blanco, J.E., Blanco, J., Garcia-Gil, L.J., and Darfeuille-Michaud, A. (2009). Molecular diversity of Escherichia coli in the human gut: new ecological evidence supporting the role of adherent-invasive E. coli (AIEC) in Crohn’s disease. Inflamm. Bowel Dis. 15, 872–882. https://doi.org/10.1002/ibd.20860.

20. Jacobi, C.A., and Malfertheiner, P. (2011). Escherichia coli Nissle 1917 (Metaflor): new insights into an old probiotic bacterium. Dig. Dis. 29, 600–607. https://doi.org/10.1159/000333307.

21. Sonnenborn, U. (2016). Escherichia coli strain Nissle 1917—from bench to bedside and back: history of a special Escherichia coli strain with probiotic properties. FEMS Microbiol. Lett. 363, ftw212. https://doi.org/10.1093/femsle/ftw212.

22. Sun, J., Gunzer, F., Westendorf, A.M., Buer, J., Scharfe, M., Jarek, M., Gossling, F., Blöcker, H., and Zeng, A.P. (2005). Genomic peculiarity of coding sequences and metabolic potential of probiotic Escherichia coli strain Nissle 1917 inferred from raw genome data. J. Biotechnol. 117, 147–161.
23. Kim, D., Kim, Y., and Yoon, S.H. (2021). Development of a genome-scale metabolic model and phenotype analysis of the probiotic Escherichia coli strain Nissle 1917. Int. J. Mol. Sci. 22, 2122. https://doi.org/10.3390/ijms22042122.

24. Reister, M., Hoffmeier, K., Kreizdorn, N., Rotter, B., Liang, C., Rund, S., Dandekar, T., Sonnenborn, U., and Oelschlaeger, T.A. (2014). Complete genome sequence of the Gram-negative probiotic Escherichia coli strain Nissle 1917. J. Bacteriol. 187, 106–107. https://doi.org/10.1128/JB.00442-14.

25. Hayashi, K., Morooka, N., Yamamoto, Y., Fujita, K., Isomo, K., Choi, S., Ohtsubo, E., Baba, T., Wanner, B.L., Mori, H., and Horuchi, T. (2006). Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110. Mol. Syst. Biol. 2, 2006.0007. https://doi.org/10.1038/msb4100049.

26. Brinkkötter, A., Klüss, H., Alpert, C., and Lengeler, J.W. (2000). Pathways for the utilization of N-acetyl-galactosamine and galactosamine in Escherichia coli. Mol. Microbiol. 37, 125–135. https://doi.org/10.1046/j.1365-2958.2000.01969.x.

27. Lengeler, J. (1977). Analysis of mutations affecting the dissimilation of galactitol (dulcitol) in Escherichia coli K12. Mol. Gen. Genet. 152, 83–91. https://doi.org/10.1007/BF00264494.

28. Dalgaard, F. (1983). Molecular cloning of the gene for phosphofructokinase-2 of Escherichia coli and the nature of a mutation, pF81, causing a high level of the enzyme. J. Mol. Biol. 168, 285–305. https://doi.org/10.1016/0022-2836(83)90019-9.

29. Brinkkötter, A., Shakeri-Garakani, A., and Lengeler, J.W. (2002). Two class II pfkB1 regulators of Escherichia coli K-12 strains MG1655 and W3110. Mol. Gen. Genet. 162, 410–419. https://doi.org/10.1007/s00294-002-0040-6.

30. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Matsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. 2, 2006.0008. https://doi.org/10.1038/msb4100050.

31. Cutito-Jiménez, A.M., Martins-Pinheiro, M., Lima, W.C., Martin-Tornet, A., Morales, O.G., and Menck, C.F.M. (2010). Evolutionary placement of Xanthomonadales based on conserved protein signature sequences. Mol. Phylogenet. Evol. 54, 524–534. https://doi.org/10.1016/j.ympev.2009.09.026.

32. Ludwig, W., Schleifer, K.-H., and Whitman, W.B. (2009). Revised road map to the phylog Firmicutes. In Bergey’s Manual of Systematic Bacteriology, P. De Vos, G.M. Garrity, D. Linskens, K.-H. Schleifer, and W.B. Whitman, eds. (Springer), pp. 1–13. https://doi.org/10.1007/978-0-387-68489-5_1.

33. Kang, H., Xu, X., Fu, K., An, X., Mi, Z., Yin, X., Peng, F., Pei, G., Wang, Y., Huang, Y., et al. (2015). Characterization and genomic analysis of quinolone-resistant Delftia sp. 670 isolated from a patient who died from severe pneumonia. Curr. Microbiol. 71, S4–61. https://doi.org/10.1007/s00284-015-0818-6.

34. Hördt, A., López, M.G., Meier-Kolthoff, J.P., Schlegel, M., Wehnold, L.M., Tindall, B.J., Gronow, S., Kyrpides, N.C., Wayne, T., and Goker, M. (2020). Analysis of 1,000+ type-strain genomes substantially improves taxonomic classification of Alphaproteobacteria. Front. Microbiol. 11, 468. https://doi.org/10.3389/fmicb.2020.00468.

35. Ashford, R.T., Muchowski, J., Koylass, M., Scholz, H.C., and Whatmore, A.M. (2020). Application of whole genome sequencing and pan-family multi-locus sequence analysis to characterize relationships within the family Brucellaceae. Front. Microbiol. 11, 1329. https://doi.org/10.3389/fmicb.2020.01329.

36. Baumler, D.J., Ma, B., Reed, J.L., and Perna, N.T. (2013). Inheriting ancient metabolism using ancestral core metabolic models of enterobacteria. BMC Syst. Biol. 7, 46. https://doi.org/10.1186/1752-0509-7-46.

37. Gao, B., Mohan, R., and Gupta, R.S. (2009). Phylogenomics and protein signatures elucidating the evolutionary relationships among the Gammaproteobacteria. Int. J. Syst. Evol. Microbiol. 59, 234–247. https://doi.org/10.1099/ijs.0.00741-0.

38. Guo, Y., Takashima, Y., Sato, Y., Narisawa, K., Ohta, H., and Nishizawa, T. (2020). Mycoavidus sp. strain B2-EB: comparative genomics reveals minimal genomic features required by a cultivable Burkholderiaceae-related endo fungal bacterium. Appl. Environ. Microbiol. 86, e01018-20. https://doi.org/10.1128/AEM.01918-20.

39. Munoz-Gomez, S.A., Hess, S., Burger, G., Lang, B.F., Susko, E., Slamovits, C.H., and Roger, A.J. (2019). An updated phylogeny of the Alphaproteobacteria reveals that the parasitic Rickettsiales and Holosporales have independent origins. Elle & eLife 8, e42535. https://doi.org/10.7554/eLife.42535.

40. Rahi, P., Khairnar, M., and Palsson, B.O. (2018). Comparative genomics of transport proteins in probiotic and pathogenic Escherichia coli and Salmonella enterica strains. Microb. Pathol. 107, 106–115. https://doi.org/10.1016/j.micpath.2017.03.022.
52. Wichelecki, D.J., Vetting, M.W., Chou, L., Al-Obaidi, N., Bouvier, J.T., Almo, S.C., and Gerlt, J.A. (2015). ATP-binding Cassette (ABC) transport system solute-binding protein-guided identification of novel D-altitol and galactitol catabolic pathways in Agrobacterium tumefaciens C58. J. Biol. Chem. 290, 28963–28976. https://doi.org/10.1074/jbc.M115.686857.

53. David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., et al. (2014). Diet rapidly and reproducibly alters the human gut microbiome. Nature 505, 559–563. https://doi.org/10.1038/nature12820.

54. Suez, J., Korem, T., Zeevi, D., Zilberman-Schapira, G., Thaiss, C.A., Mazo, O., Israeli, D., Zmora, N., Gilad, S., Weinberger, A., et al. (2014). Artificial sweeteners induce glucose intolerance by altering the gut microbiota. Nature 514, 181–186. https://doi.org/10.1038/nature13793.

55. Poo, H., Song, J.J., Hong, S.-P., Choi, Y.-H., Yun, S.W., Kim, J.-H., Lee, S.C., Lee, S.-G., and Sung, M.H. (2002). Novel high-level constitutive expression system, pHCE vector, for a convenient and cost-effective soluble production of human tumor necrosis factor-α. Biotechnol. Lett. 24, 1185–1189. https://doi.org/10.1023/A:1016107230825.

56. Palmeros, B., Wild, J., Szybalski, W., Le Borgne, S., Hernández-Chávez, G., Gosset, G., Valle, F., and Bolívar, F. (2000). A family of removable cassettes designed to obtain antibiotic-resistance-free genomic modifications of Escherichia coli and other bacteria. Gene 247, 255–264. https://doi.org/10.1016/s0378-1119(00)00075-5.

57. Kim, H., Kim, S., Kim, D., and Yoon, S.H. (2020). A single amino acid substitution in aromatic hydroxylase (HpaB) of Escherichia coli alters substrate specificity of the structural isomers of hydroxyphenylacetate. BMC Microbiol. 20, 109. https://doi.org/10.1186/s12866-020-01798-4.

58. Kim, S., Jeong, H., Kim, E.Y., Kim, J.F., Lee, S.Y., and Yoon, S.H. (2017). Genomic and transcriptomic landscape of Escherichia coli BL21(DE3). Nucleic Acids Res. 45, S285–S293. https://doi.org/10.1093/nar/gkx228.

59. Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCt Method. Methods 25, 402–408. https://doi.org/10.1006/meth.2001.1262.
### STAR METHODS

**KEY RESOURCES TABLE**

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| **Bacterial strains** |        |            |
| *Escherichia coli* Nissle 1917 | Ardeypharm GmbH | Referred to as EcN |
| EcNΔ1 | This paper | Deletion of ECOLIN_RS18675 in EcN |
| EcNΔ6 | This paper | Deletion of ECOLIN_RS18670-RS18695 in EcN |
| EcNΔ6/pHCE-6 | This paper | EcNΔ6 with pHCE-6 |
| *E. coli* BL21(DE3) | Lab stock | N/A |
| *E. coli* W | Lab stock | N/A |
| *E. coli* MG1655 | Lab stock | N/A |
| *E. coli* BW25113 | Lab stock | N/A |
| BL21(DE3)/pHCE-6 | This paper | *E. coli* BL21(DE3) with pHCE-6 |
| W/pHCE-6 | This paper | *E. coli* W with pHCE-6 |
| MG1655/pHCE-6 | This paper | *E. coli* MG1655 with pHCE-6 |
| BW25113/pHCE-6 | This paper | *E. coli* BW25113 with pHCE-6 |
| BW25113ΔpfkA | Keio collection | Cat# JW3887-KC |
| BW25113ΔpfkB | Keio collection | Cat# JW5280-KC |
| BW25113ΔpfkA/pHCE-6 | This paper | BW25113ΔpfkA with pHCE-6 |
| BW25113ΔpfkB/pHCE-6 | This paper | BW25113ΔpfkB with pHCE-6 |
| **Deposited data** |        |            |
| *Escherichia coli* Nissle 1917 reference genome | RefSeq | Accession number: NZ_CP007799 |
| **Oligonucleotides** |        |            |
| See Table S3 for primers used for gene deletion mutant construction, cloning, and quantitative real-time PCR | This paper | N/A |
| **Recombinant DNA** |        |            |
| pHCE-IIB | Poo et al. | High-copy-number expression vector with a constitutive promoter (HCE), Ap<sup>+</sup> |
| pHCE-6 | This paper | pHCE-IIB carrying ECOLIN_RS18670-RS18695, Ap<sup>+</sup> |
| pRedET | Gene Bridges | Cat# K006 |
| pJW168 | Palmeros et al. | Cre recombinase expression vector with an IPTG-inducible lacUV5 promoter, Ap<sup>+</sup> |
| **Software and algorithms** |        |            |
| BLAST+ 2.9.0 | NCBI website | https://ftp.ncbi.nlm.nih.gov/blast/ executables/blast+/2.9.0/ |
| ClermonTyping | Beghain et al. | https://github.com/A-BN/ClermonTyping |

### RESOURCE AVAILABILITY

**Lead contact**
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact Sung Ho Yoon (syoon@konkuk.ac.kr).

**Materials availability**
This study did not generate new unique reagent.
Data and code availability
- All data reported in this paper will be shared by the lead contact upon request.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS
Bacterial strains
The E. coli strains used in this study are listed in key resources table. E. coli strain Nissle 1917 (Mutaflor, DSM 6601, serotype O6:K5:H1) was kindly provided by Ardeypharm GmbH (Herdecke, Germany).

METHOD DETAILS
Cell culture
Aerobic cultures were performed using M9 minimal medium. The M9 medium (pH 7.0) contained 6.78 g/L Na2HPO4, 3 g/L KH2PO4, 1 g/L NH4CL, and 0.5 g/L NaCl supplemented with 0.8 g/L MgSO4·7H2O and 0.5 mL/L trace metal solution. The trace metal solution contained 0.5 g/L MnSO4·4H2O, 0.02 g/L Na2B4O7·10H2O, 2.2 g/L ZnSO4·7H2O, 1 g/L CuSO4·5H2O, 2 g/L CaCl2, 0.1 g/L (NH4)6Mo7O24·4H2O, 0.5M HCl, and 10 g/L FeSO4·7H2O. As the sole carbon source, 40mM D-tagatose or 3 g/L glucose was added to the M9 medium for bacterial growth on D-tagatose and glucose, respectively. Seed cultures were prepared by growing cells in M9 medium supplemented with glucose at 200 rpm for 12 h. Then, the seed cultures were 20-fold diluted in fresh M9 medium supplemented with D-tagatose. The main cultures were performed using a 250 mL flask containing 50 mL culture medium or a 96-well Epoch 2 Microplate Spectrophotometer (BioTek, Winooski, VT) containing 150 μL culture medium in each well. For cultures of cells harboring pHCE-6, 100 μg/mL ampicillin was added to the culture medium. All cell cultures were performed at 37°C.

For the anaerobic culture, 5 mL of LB medium (10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl) was inoculated from a frozen stock and incubated anaerobically for 5 h. Next, 0.25 mL of the aerobic culture was transferred into 5 mL LB medium, and the seed culture was grown anaerobically for 48 h. For the main culture, 5 mL M9 medium supplemented with 40 mM D-tagatose and 2 g/L yeast extract was inoculated with 0.25 mL of the anaerobic seed culture, and was incubated anaerobically for 120 h. As controls, cells were cultured using M9 medium supplemented only with 2 g/L yeast extract. To ensure strictly anaerobic condition in the seed culture and the main culture, the inoculations were performed in an anaerobic chamber (Coy Laboratory Products, Grass Lake, MI) with oxygen-free atmosphere (85% N2, 10% CO2, and 5% H2), and cells were grown in a screw-cap culture tube (MTC Bio DuoClick™, Sayreville, NJ) placed in an anaerobe container (GasPak™ EZ Anaerobe Container System, BD, Franklin Lakes, NJ). All cell cultures were performed with shaking at 200 rpm and 37°C.

Analytical procedures
Bacterial growth was monitored by measuring the optical density at 600 nm (OD600). D-Tagatose in the culture media was quantified using a 1260 Infinity HPLC system (Agilent Technologies, Santa Clara, CA) equipped with a refractive index detector (RID) and a Shodex SUGAR SP0810 column (300 × 8.0 mm; Showa Denko, Tokyo, Japan). The concentrations of glucose, acetate, citrate, formate, and succinate in the culture media were measured using an HPLC equipped with an ultraviolet (UV) detector and an Aminex HPX-87H column (300 × 7.8 mm; Bio-Rad Laboratories, Hercules, CA).

Construction of gene deletion mutants
All primers used in this study are listed in Table S3. A single gene (ECOLIN_RS18675) or gene cluster (ECOLIN_RS18670 to ECOLIN_RS18695) in EcN was deleted by the Cre/loxP recombination system, resulting in the construction of EcNΔ1 and EcNΔ6 mutants, respectively. Plasmid pRedET was used to replace the gene(s) in the EcN chromosome with a chloramphenicol-resistant gene (cat). Wild-type EcN cells were transformed with the pRedET plasmid and cultivated at 37°C. Homology-directed repair fragments were PCR amplified from cat gene-harboring plasmids with KO_6F/R primers for EcNΔ6 construction and KO_1F/R primers for EcNΔ1 construction, producing amplificons including the cat gene.
and homology arms. The PCR products were transformed into electrocompetent EcN harboring pRedET. After induction with L-arabinose (1 mM) at the OD_{600} of 0.3 for the expression of λ-red recombinases, colonies were selected on LB agar plates containing chloramphenicol (25 μg/mL). Gene replacement with the cat gene was confirmed by PCR using KO_confirmed_6F/R primers for EcNΔ6 and KO_confirmed_1F/R primers for EcNΔ1. The cat gene was removed using pJW168 expressing Cre recombinase under the control of an IPTG-inducible promoter. IPTG was added at a final concentration of 1 mM at OD_{600} of 0.5.

Construction of an expression vector
The gene cluster was PCR-amplified from the genomic DNA of EcN using Exp_6_F/R primers. Plasmid pHCE-IIB was used for the constitutive expression of target genes. pHCE-IIB has a strong constitutive promoter cloned from the thermostable D-amino acid aminotransferase (D-AAT) gene of Geobacillus toebii.\textsuperscript{55} pHCE-IIB was digested with BamHI and XbaI restriction enzymes, and then inserted with the amplified DNA fragment using an EZ-Fusion\textsuperscript{TM} HT Cloning kit (Enzynomics, Seoul, South Korea). The constructed plasmid (pHCE-6) was electroporated into E. coli cells.

Quantitative real-time PCR (qRT-PCR)
Wild-type EcN was grown in M9 minimal medium supplemented with 3 g/L glucose or 40 mM D-tagatose as the sole carbon source. Total RNA sampled at OD_{600} of 1.0, was extracted using the mirVana miRNA Isolation Kit (ThermoFisher Scientific, Waltham, MA) according to the manufacturer’s instructions. cDNA was amplified using the ReverTra Ace qPCR RT Master Mix with the gDNA Remover Kit (Toyobo, Japan). qRT-PCR was performed using a QuantStudio\textsuperscript{®} 3 Real-Time PCR System (Applied Biosystems; Thermo Fisher Scientific, Inc., MA). Each qRT-PCR reaction contained 1 μL of diluted cDNA, 10 pmol of each primer (Table S3), and 10 μL of iQ\textsuperscript{TM} SYBR Green Supermix (Bio-Rad, Hercules, CA) and was performed on a 96-well plate for 40 cycles. The data were analyzed using QuantStudio\textsuperscript{TM} Design & Analysis Software v1.4.3 (Applied Biosystems). The 5S rRNA gene was used as an endogenous control for normalization of gene expression. The relative gene expression level in reference to cells grown on glucose was calculated using the 2^{−ΔΔCt} method.\textsuperscript{59} The reference samples (glucose growth) were calculated as 2^{−ΔΔCt} values of 1.

Homology search
As of August 2021, proteome files of 24,025 complete bacterial genomes were downloaded from the NCBI GenBank database (Table S1). Each of the eight ORFs in the gene clusters of EcN was used as the query in BLASTP searches against the set of protein sequences from each of the downloaded proteomes. If the identity of the resulting best hit was over 30% and the aligned region was both over 70% of the lengths of the query and the hit, the pair of sequences was considered as a homolog. A genomic region containing at least four homologs of tagatose kinase (ECOLIN_RS18675) and ABC transport subunits (ECOLIN_RS18680, ECOLIN_RS18685, and ECOLIN_RS18690) was identified in each genome.

A list of 110 E. coli strains, including 53 isolates from IBD patients and healthysubjects, was obtained from a previous study.\textsuperscript{49} Their genome sequences (97 complete and 13 incomplete) were downloaded from NCBI GenBank and Australian National University (ANU) (Table S2). Homologous regions were searched using BLASTN against 110 E. coli listed in a previous study.\textsuperscript{49} The pair of sequences was considered a homolog if the aligned region was both over 90% of the lengths of the query and the hit. The resulting best hit was 93.7% at the minimum, and homologous regions were manually inspected.

QUANTIFICATION AND STATISTICAL ANALYSIS
For qRT-PCR analysis, Welch’s t-test was used to evaluate the differences between the two groups. p-values <0.05 were considered statistically significant (* p-value <0.05; ** p-value <0.01). Data were expressed as mean ± the standard error of the mean (SEM).