

RESEARCH

Open Access



Regulation of the tagatose catabolic gene cluster and development of a tagatose-inducible expression system in the probiotic *Escherichia coli* Nissle 1917

Keunha Park^{1†}, Youngshin Kim^{2†}, Dohyeon Kim², Su Min Oh¹, Tae Jeong Koo¹, Seung Min Yoo^{1*} and Sung Ho Yoon^{2*}

Abstract

Background The probiotic *Escherichia coli* Nissle 1917 (EcN) is a promising microbial chassis for therapeutic and industrial applications. However, its broad utility is limited by a lack of reliable inducible gene expression systems that precisely control gene expression.

Results We developed a tagatose-inducible expression system in EcN using D-tagatose, a naturally occurring sugar with established safety in humans, as a metabolizable inducer. Through differential RNA sequencing and sequence analysis, we identified the key regulatory elements governing D-tagatose catabolism in EcN and demonstrated that the DeoR family regulator (TagR) functions as a tagatose-responsive repressor. The developed system exhibited a strong dose-dependent response to D-tagatose, ensuring uniform and tunable gene activation across cell populations. Additionally, a catabolite repression-enabled auto-induction strategy facilitated robust biomass accumulation, followed by targeted protein production. This expression system was successfully applied to overexpress recombinant proteins under both aerobic and anaerobic conditions.

Conclusions D-Tagatose is a naturally occurring low-calorie sugar that can serve as an inducer in vivo, including within the human gut microbiome. Thus, the tagatose-inducible expression system provides EcN with an additional tunable option for gene regulation, which may be valuable in applications such as synthetic biology and metabolic engineering.

Keywords *Escherichia coli* Nissle 1917, Probiotics, Tagatose, Inducible gene expression system, Tagatose catabolic gene cluster, Regulatory elements

[†]Keunha Park and Youngshin Kim contributed equally to this work.

*Correspondence:

Seung Min Yoo
yooseun1@cau.ac.kr
Sung Ho Yoon
syoon@konkuk.ac.kr

¹School of Integrative Engineering, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Republic of Korea

²Department of Bioscience and Biotechnology, Konkuk University, Seoul 05029, Republic of Korea



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

Genetic engineering of probiotics such as *Lactobacillus* and *Bifidobacterium* species has great potential as an innovative therapeutic approach because of their extensive metabolic potential and essential roles in host health and disease [1, 2]. Among these, *Escherichia coli* Nissle 1917 (EcN), a gram-negative probiotic bacterium, has a well-established safety profile and has long been used to treat intestinal disorders, such as inflammatory bowel disease (IBD), ulcerative colitis, and diarrhea [3, 4]. With its probiotic properties, proven safety record, and amenability to genetic engineering, EcN has emerged as a versatile platform for developing living therapeutics to diagnose, prevent, and treat a variety of diseases [5, 6]. Notably, EcN has been genetically engineered as a living therapeutic strategy for colorectal cancer, IBD, and other metabolic diseases [5].

The development of live bacterial therapeutics hinges on precisely modulating gene expression so that the production of therapeutic molecules is restricted to the intended tissues or cells, thereby minimizing off-target effects and preventing harm to normal cells. Achieving targeted expression relies on the use of robust inducible promoters capable of ensuring high-level protein or metabolite production in specific environments. In this context, EcN has been engineered by inserting the T7 RNA polymerase gene into its chromosome under the control of the lacUV5 promoter, which is activated by the addition of isopropyl- β -D-1-thiogalactopyranoside (IPTG) to the growth medium [7–9]. However, the T7 system often competes with housekeeping protein expression for RNA synthesis resources, causing a metabolic burden and reducing bacterial growth [10].

Consequently, circumventing this toxicity typically requires the complex optimization of target protein expression [11], potentially limiting the broader applicability of the expression system. An alternative to chemical induction involves the use of light to induce gene expression, which offers a less toxic and more convenient solution in clinical settings. Light-inducible gene expression systems have been explored in engineered EcN to produce anti-tumor proteins within tumors in response to near-infrared (NIR) light [12, 13]. However, the limited penetration of NIR light through the body poses challenges for its clinical application [5]. Thus, the development of alternative, clinically feasible induction methods may further enhance the regulation of therapeutic protein expression in EcN for medical and industrial applications.

In clinical applications, the choice of an inducer for gene expression systems prioritizes nontoxic and biocompatible options to ensure patient safety. Although

synthetic compounds such as IPTG are effective in laboratory settings, they pose potential health risks owing to their possible chronic toxicity and uncertain long-term effects. Alternatively, naturally occurring sugars, such as arabinose and xylose, are appealing for clinical applications because of their low cost and biocompatibility. However, their presence in the human diet and rapid metabolism in the human body can lead to inconsistent inducer levels, limiting their ability to reliably activate engineered bacteria in the gut. Consequently, it is necessary to investigate naturally occurring sugars that are rare in the human diet and are not synthesized by the human body yet remain biocompatible and effective as inducers.

D-Tagatose, commonly called tagatose, is a rare, low-calorie sugar naturally present in small amounts in dairy products and fruits [14]. Recognized as safe by the U.S. Food and Drug Administration [15], it is widely used as a sweetener in food and beverages. As most ingested D-tagatose is poorly digested by human metabolism, it is predominantly catabolized by a limited number of enteric microorganisms, primarily lactic acid bacteria such as *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Enterococcus*, *Pediococcus*, and *Streptococcus* [16]. Previously, we identified a gene cluster responsible for tagatose utilization in EcN, challenging the common perception that *E. coli* strains cannot metabolize D-tagatose [17]. In the present study, we investigated the regulatory elements within this gene cluster and constructed a tagatose-inducible expression vector to develop a tunable expression system for engineering EcN. We evaluated the performance of this expression system and successfully used it to overproduce two recombinant proteins under both aerobic and anaerobic culture conditions.

Methods

Bacterial strains and cell culture

The *E. coli* strains and plasmids used in this study are listed in Table S1. *E. coli* strain Nissle 1917 (Mutaflor, DSM6601, serotype O6:K5:H1) was provided by Ardeypharm GmbH (Herdecke, Germany). For aerobic cultures, an overnight culture was prepared by inoculating 5 mL of M9 minimal medium supplemented with 3 g/L glucose from frozen cell stock. The seed culture (5 mL) was then transferred into a 250 mL baffled flask containing 25 mL M9 medium supplemented with D-tagatose and/or glucose. The M9 minimal medium (pH 7.0) contained per liter: 6.78 g Na_2HPO_4 , 3 g KH_2PO_4 , 1 g NH_4Cl , and 0.5 g NaCl , 0.8 g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, and 0.5 mL of trace metal solution. The trace metal solution contains 10 g/L $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 2.25 g/L $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 1 g/L $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 0.5 g/L of $\text{MnSO}_4 \cdot 5\text{H}_2\text{O}$, 0.23 g/L of $\text{Na}_2\text{B}_4\text{O}_7 \cdot 10\text{H}_2\text{O}$, 2 g/L of

$\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 0.1 g/L of $(\text{NH}_4)_6\text{Mo}_7\text{O}_{24}$, and 5 M HCl. For anaerobic cultures, an initial seed culture was prepared by inoculating 5 mL of LB medium (containing 10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl) from a frozen stock and incubating aerobically for 5 h. Subsequently, 0.25 mL of the culture was transferred into 5 mL of fresh LB medium and grown anaerobically for 48 h. For the main anaerobic culture, 0.25 mL of the anaerobic seed culture was inoculated into 5 mL of M9 medium supplemented with 2 g/L yeast extract and D-tagatose and/or glucose. The cultures were grown anaerobically for 72 h. To maintain strict anaerobic culture conditions, the inoculations were performed in an anaerobic chamber (Coy Laboratory Products, Grass Lake, MI) with an oxygen-free atmosphere (85% N_2 , 10% CO_2 , and 5% H_2). Cells were cultivated in screw-cap culture tubes (MTC Bio DuoClick™, Sayreville, NJ) placed in an anaerobe container (GasPak™ EZ Anaerobe Container System, BD, Franklin Lakes, NJ). All cultures were supplemented with 100 µg/mL ampicillin and incubated at 37 °C with shaking at 200 rpm.

The concentrations of glucose and D-tagatose in the culture media were measured using high-performance liquid chromatography (HPLC; 1260 Infinity II, Agilent Technologies) equipped with a Shodex SUGAR SP0810 column (300 × 8.0 mm; Showa Denko, Tokyo, Japan). Culture samples were centrifuged at 2,616 ×g for 3 min, and the supernatant was filtered through a 0.2 µm syringe filter (Nylon syringe filter, HM). The mobile phase (100% water) was delivered at a flow rate of 1 mL/min, and sugar concentrations were detected using a refractive index detector (G7162A; Agilent Technologies).

Construction of expression plasmids

The primer sequences are listed in Table S2. *E. coli* DH5α was used for gene manipulation. The pCop-mCherry plasmid [18] was used as a backbone to construct the expression plasmids, pCop- R_{int} -mCherry, pCop- $R_{\text{int-tagR}}$ -mCherry, and pCop- P_{J23105} -mCherry, by replacing its original P_L promoter at the SacI and PacI sites with R_{int} , $R_{\text{int-tagR}}$, or P_{J23105} , respectively. The R_{int} and $R_{\text{int-tagR}}$ regions were amplified from EcN genomic DNA using the primer pairs $R_{\text{int-F_SacI}}/R_{\text{int-R_PacI}}$ and $R_{\text{int-tagR-F_SacI}}/R_{\text{int-tagR-R_PacI}}$, respectively. The P_{J23105} promoter fragment was obtained by digesting the p15C- P_{J23105} -Hfq(R16A) plasmid [18] with SacI and PacI. Each fragment was then inserted into pCop-mCherry at the SacI and PacI sites, generating the corresponding expression plasmids. The complete sequences and annotations of pCop- R_{int} -mCherry and pCop- $R_{\text{int-tagR}}$ -mCherry, and pCop- P_{J23105} -mCherry

are provided in the supplementary material as a GenBank-formatted file ([pCop-X-mCherry.gb](https://doi.org/10.26434/chemrxiv-2024-12345)).

MBP-hEGF and polyethylene terephthalate hydrolase (PETase) expression plasmids were constructed using pCop- R_{int} -mCherry as the backbone. The human epidermal growth factor (hEGF) gene was fused to the maltose-binding protein (MBP) gene, and a His-tag was included for western blot analysis [19]. The MBP-hEGF-His fusion protein gene was PCR amplified using the primer pair hEGF_F_PacI/hEGF_R_PstI and cloned into the PacI and PstI sites of pCop- R_{int} -mCherry, generating pCop- R_{int} -MBP-hEGF. PETase was fused to the N-terminal Lamb signal sequence (Lamb_{ss}), and a His-tag was included for western blot analysis. The Lamb_{ss} -PETase-His fusion protein gene was amplified using hEGF_F_PacI/PETase_R_SpeI primers and cloned into the PacI and SpeI sites of pCop- R_{int} -mCherry to generate pCop- R_{int} -PETase. Plasmid maps of the expression constructs developed in this study are shown in Fig. S1.

Protein detection

For mCherry fluorescence measurements, the cells were centrifuged at 2,616 ×g for 3 min, and the cell pellets were washed and resuspended in PBS buffer (pH 8.0). The mCherry fluorescence intensity (excitation at 587 nm and emission at 601 nm) was quantified using flow cytometry (easyCyte, Luminex, Austin, TX, USA).

Western blotting was performed to measure MBP-hEGF and PETase protein expression. Whole intracellular proteins were electrophoresed using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Cells were harvested by centrifugation at 2,616 ×g for 3 min, and the cell pellet was resuspended in SDS-PAGE sample buffer [200 mM Tris/HCl (pH 6.8), 30% (v/v) glycerol, 8% (w/v) SDS, 0.1% (w/v) Orange G]. The mixed samples were heated at 98 °C for 10 min to prepare the whole protein lysates. Proteins were separated by electrophoresis on a 12% acrylamide gel, followed by staining with Coomassie blue. For western blotting, proteins from the SDS-PAGE gel were transferred onto a nitrocellulose membrane (XOGENE, Englewood, NJ, USA) at 200 mA for 1.5 h. The membrane was probed with a monoclonal mouse 6× His epitope-tag antibody (1:1000; Invitrogen, Carlsbad, CA, USA). After three washes with Tris-buffered saline-Tween 20 for 10 min each, a goat anti-mouse horseradish peroxidase-conjugated secondary antibody (1:5000; Invitrogen) was applied to the membrane and incubated at room temperature for 1 h. Bound conjugates were detected by adding 3,3',5,5'-tetramethylbenzidine (TMB) blotting solution (Thermo Scientific, Waltham, MA, USA). Band

intensities were quantified using ImageJ software (<https://imagej.net/ij/>) [20] and normalized to the highest target protein expression set at 1.

Differential RNA sequencing (dRNA-seq)

Total RNA was extracted using the RNeasy Protect Bacteria Reagent (Qiagen, Düsseldorf, Germany) and mirVana miRNA Isolation Kit (Thermo Fisher Scientific Inc., Waltham, MA, USA), as described previously [21]. The dRNA-seq library was constructed according to a previously published protocol [22] with modifications. To distinguish primary transcripts from processed RNAs, total RNA was treated with terminator 5'-phosphate-dependent exonuclease (TEX), which selectively degrades processed RNAs while preserving primary RNAs. Two RNA libraries were prepared: one from TEX-treated RNA (enriched in primary transcripts) and the other from untreated RNA (containing both primary and processed transcripts). A poly(A) tail was added to the 3' end of the RNA molecules using poly(A) polymerase, and 5'-triphosphate residues were converted to 5'-monophosphate using tobacco acid pyrophosphatase. RNA adapters were then ligated to the 5' ends of RNA fragments. First-strand cDNA was synthesized using oligo(dT) adapter primers and Moloney murine leukemia virus reverse transcriptase, followed by PCR amplification. The resulting cDNA libraries were sequenced using an Illumina NextSeq 500 system (Illumina, San Diego, CA, USA) with 75 bp single-end reads.

The quality of the raw RNA sequencing data was assessed using FastQC (<https://qubeshub.org/resources/fastqc>). The 5'-adaptor sequences were trimmed using Trimmomatic [23], and the reads were aligned to the EcN genome (RefSeq accession number of chromosome: NZ_CP058217.1, pMUT1: NZ_CP058218.1, and pMUT2: NZ_CP058219.1) using READemption [24]. To account for differences in the total number of reads across libraries, the number of mapped reads per nucleotide was normalized to reads per million (RPM). Transcription start sites (TSSs) were identified using ANNOgesic [25], and the read depth at each genome coordinate was visualized using the Integrated Genome Browser [26].

Analysis of regulatory elements

The promoter regions within the *tagR*–*tagY* intergenic region were predicted by analyzing sequences ranging from +21 to –60 bp relative to each TSS. iPromoter-2 L [27] was used to predict the σ factors suitable for each promoter. Both P_{tag} and P_{reg} were predicted to be recognized by the σ^{70} factor, prompting further analysis of their –10 and –35 regions using BPROM [28], a tool specifically designed for bacterial

σ^{70} promoter prediction. The predicted sequences of the –10 and –35 regions were revised based on σ^{70} consensus sequences. CRPBSFinder [29] and Pred-CRP [30] were used to predict the binding sites and regulatory rules of the cyclic AMP receptor protein (CRP). The binding sites of TagR were identified using TFbinding [31]. The predicted binding sites were further refined through multiple sequence alignments using Clustal Omega [32, 33].

Results and discussion

Identification of regulatory elements within the *tag* locus

Previously, we identified an 8.5 kb genomic region in EcN that is absent in *E. coli* K-12 [17]. This region consists of eight genes organized into two outwardly oriented clusters. The first cluster, hereafter referred to as the *tag* cluster, contains six genes encoding D-tagatose biphosphate aldolase, ATP-binding cassette transporter, D-tagatose kinase, and aldose 1-epimerase. The second cluster, termed the *reg* cluster, includes two genes encoding a DeoR family transcriptional regulator and a sugar kinase (Fig. 1A). Following the bacterial gene nomenclature [34], we named these eight genes in the EcN chromosome (RefSeq accession NZ_CP058217) as follows: *tagK* (locus tag: HW372_RS02435), *tagR* (HW372_RS02440), *tagY* (HW372_RS02445), *tagA* (HW372_RS02450), *tagB* (HW372_RS02455), *tagC* (HW372_RS02460), *tagD* (HW372_RS02465), and *tagE* (HW372_RS02470). Experimental validation has demonstrated that *tag* cluster genes are essential for D-tagatose utilization, whereas the functions of *reg* cluster genes remain unexplored [17]. When EcN was grown on D-tagatose instead of glucose, the expression of the first three genes (*tagYAB*) in the *tag* cluster increased by over 1000-fold, whereas the expression of *tagRK* in the *reg* cluster increased by approximately two-fold [17]. This dramatic difference suggests the presence of a highly specialized and efficient regulatory mechanism that tightly controls the expression of the *tag* cluster in response to D-tagatose availability.

To investigate the regulatory mechanism of the *tag* cluster, we identified key regulatory elements within the 341 bp intergenic region between the *tag* and *reg* clusters, including promoters, TSSs, and other potential regulatory sequences. dRNA-seq was performed to map the TSSs of EcN grown on D-tagatose as the sole carbon source (Fig. 1A). Two TSSs were identified: one located 64 nucleotides upstream of the start codon of *tagY* (the first gene in the *tag* cluster), and the other 11 nucleotides upstream of the start codon of *tagR* (the first gene in the *reg* cluster). No additional TSSs were detected within the 8.5 kb genomic region,

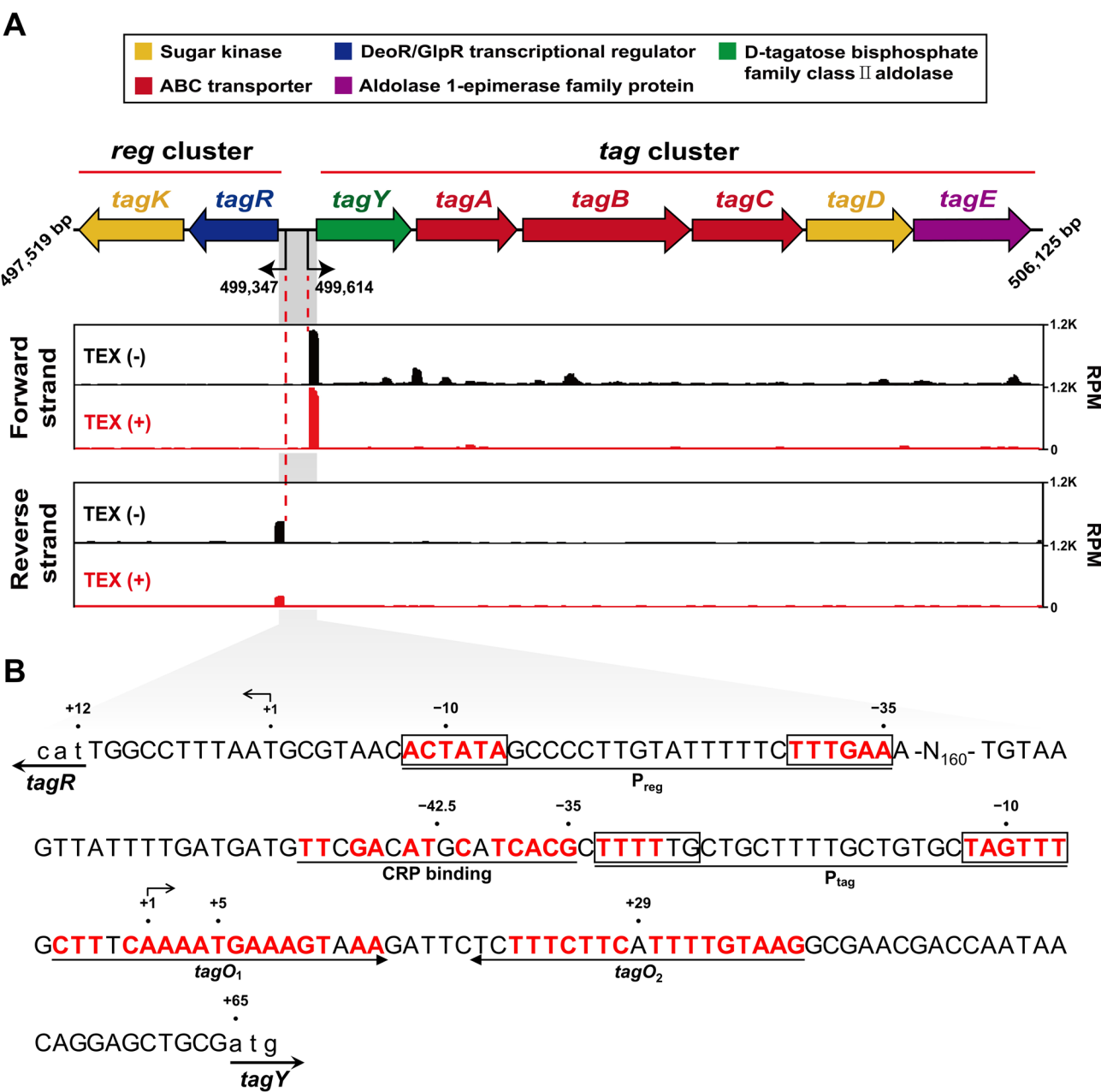


Fig. 1 Identification of regulatory elements in the *tag* locus in *E. coli* Nissle 1917 (EcN). **(A)** Organization of the *tag* and *reg* clusters with their respective transcription start sites (TSSs). The upper panel displays the genomic arrangement on the EcN chromosome (RefSeq accession number: NZ_CP058217). TSSs were identified via dRNA-seq using RNA samples treated with terminator 5'-phosphate-dependent exonuclease (TEX +), red and untreated RNA (TEX -), black. The lower panel presents normalized coverage (reads per million, RPM) mapped to strand-specific genome coordinates, with TSSs marked by red dashed lines and bent arrows indicate transcription direction. **(B)** Analysis of regulatory elements in the *tagR*–*tagY* intergenic region. The promoters (P_{tag} and P_{reg}), CRP binding site, and operator sites for TagR are indicated. The initiation codons for *tagR* and *tagY* are located at the beginning and end of the sequence, respectively. Their TSSs are positioned as +1 and marked with arrows indicating transcription direction. The predicted –10 and –35 regions of the promoters are boxed. The predicted CRP binding site is underlined, whereas the operator sites ($tagO_1$ and $tagO_2$) recognized by TagR are indicated by horizontal arrows indicating their orientation. Conserved sequences from the multiple alignment at promoters and transcription factor binding sites are highlighted in red

indicating that the two clusters were expressed as separate operons.

Sequence analysis of the *tagR*–*tagY* intergenic region revealed two promoters, designated P_{tag} and P_{reg} , which divergently direct the transcription of the *tag* cluster and *reg* cluster, respectively (Fig. 1B). The promoter regions were predicted to be located 7–33 bp upstream of the TSS for P_{tag} and 8–35 bp upstream of

the TSS for P_{reg} . These predictions were further refined through multiple sequence alignments of homologous loci resembling the 8.5 kb EcN genomic region found in *Citrobacter farmeri*, *Enterobacter hormaechei*, and *Escherichia* sp. E4742 (Fig. 2). The P_{tag} promoter contains recognizable -10 (TAGTTT) and -35 (TTTTTG) regions with 15 bp spacing, showing similarity to the consensus sequence of *rpoD* sigma factor (σ^{70}). Similarly, P_{reg} includes -10 (TATAGT) and -35 (TTCAAA) regions with 16 bp spacing, also resembling the σ^{70} consensus sequences. These observations suggest that both the P_{tag} and P_{reg} promoters are likely to be recognized and activated by the σ^{70} factors.

DeoR family regulators typically act as transcriptional repressors or activators of carbon source metabolism, often contributing to catabolite repression by promoting the preferential utilization of glucose over other carbon sources [35, 36]. Therefore, it can be inferred that the DeoR family regulator gene (*tagR*) within the *reg* cluster modulates the expression of genes in the *tag* cluster. In support of this, consensus-binding sequences for both the CRP and TagR were identified near the P_{tag} promoter (Figs. 1B and 2). The predicted CRP binding site (TTCGA- N_6 -TCACG) shares 7 out of 10 bp with the *E. coli* consensus sequence (TGTGA- N_6 -TCACA), and is located just upstream of the -35 region in the P_{tag} promoter. Two operator sequences, *tagO*₁ (CTTTCAAAATGAAAGTAAA) and *tagO*₂ (CTTACAAAATGAAGAAAGA), were predicted to be recognized by TagR. The *tagO*₁ overlaps with the TSS of the *tag* cluster, whereas *tagO*₂ is located on the opposite strand adjacent to *tagO*₁. These putative operators share 14 of 19 bp, which is consistent with the oligomeric nature of DeoR family proteins [36]. TagR is likely to undergo conformational changes upon binding to D-tagatose, thereby alleviating repression of the *tag* cluster. Interestingly, the arrangement of the predicted CRP and TagR binding sites within the *tagR*–*tagY* intergenic region resembled that of the *aga* gene cluster, which is responsible for the transport and catabolism of *N*-acetylgalactosamine and galactosamine in *E. coli* K-12 [36]. The *aga* gene cluster is negatively regulated by AgaR, a DeoR family transcriptional repressor, suggesting that a similar regulatory mechanism may operate for TagR.

Combined with the experimental findings on TagR function and catabolite repression of the *tag* cluster (as described below), the presence of these regulatory elements supports the proposed model for the regulation of the *tag* cluster (Fig. 3A). The *tag* cluster is subject to catabolite repression and is transcriptionally regulated by the tagatose-responsive repressor TagR, which is produced at a relatively constant level. In the absence of D-tagatose in the culture media, TagR binds

to the operator (*tagO*), preventing the transcription of the *tag* cluster. However, when D-tagatose is present and glucose is absent, D-tagatose acts as an inducer by binding to TagR, inducing a conformational change in TagR, which prevents it from attaching to *tagO*, thus allowing the transcription of the *tag* cluster. Although further experimental validation, such as direct biochemical confirmation of TagR binding to its operator site, is needed, the proposed regulatory mechanism should be regarded as a working hypothesis based on the available evidence, providing a basis for understanding how the *tag* genes are regulated in response to environmental D-tagatose.

Validation of TagR as a tagatose-responsive repressor of the *tag* cluster

To validate TagR as a tagatose-responsive repressor of the *tag* cluster, we constructed two pCop-based plasmids in which the mCherry fluorescent protein gene was inserted downstream of two different DNA segments: the first plasmid (pCop- R_{int} -mCherry) containing the *tagR*–*tagY* intergenic region (designated as R_{int}) and the second plasmid (pCop- $R_{int-tagR}$ -mCherry), including both the intergenic region and *tagR* gene ($R_{int-tagR}$) (Fig. 3B). The pCop backbone is a medium-copy plasmid carrying the *colE1* replicon [18]. As a reference, we used a plasmid containing the constitutive P_{J23105} promoter (pCop- P_{J23105} -mCherry) lacking the CRP-binding site.

Each construct was introduced into EcN cells and cultured in minimal medium containing either glucose or D-tagatose as the sole carbon source (Fig. 4A). In cells carrying pCop- P_{J23105} -mCherry, mCherry expression remained low, and no growth repression was observed under either glucose or D-tagatose conditions. By contrast, cells harboring pCop- $R_{int-tagR}$ -mCherry exhibited nearly complete repression of mCherry expression under both carbon sources, but exhibited severe growth repression only under D-tagatose conditions. Overexpression of TagR from the plasmid likely exceeded the available D-tagatose, preventing the effective induction of mCherry expression. Additionally, excess TagR appears to repress the chromosomal *tag* cluster, which is essential for D-tagatose utilization, thereby impairing cell growth on D-tagatose.

Supporting this, the pCop- R_{int} -mCherry construct, which included R_{int} but lacked the *tagR* gene, showed substantial mCherry fluorescence in the presence of D-tagatose while maintaining robust growth. Fluorescence increased more than 10-fold after 48 h of cultivation in D-tagatose compared with that in glucose. This result demonstrated that TagR functions as a tagatose-responsive repressor of the *tag* cluster.

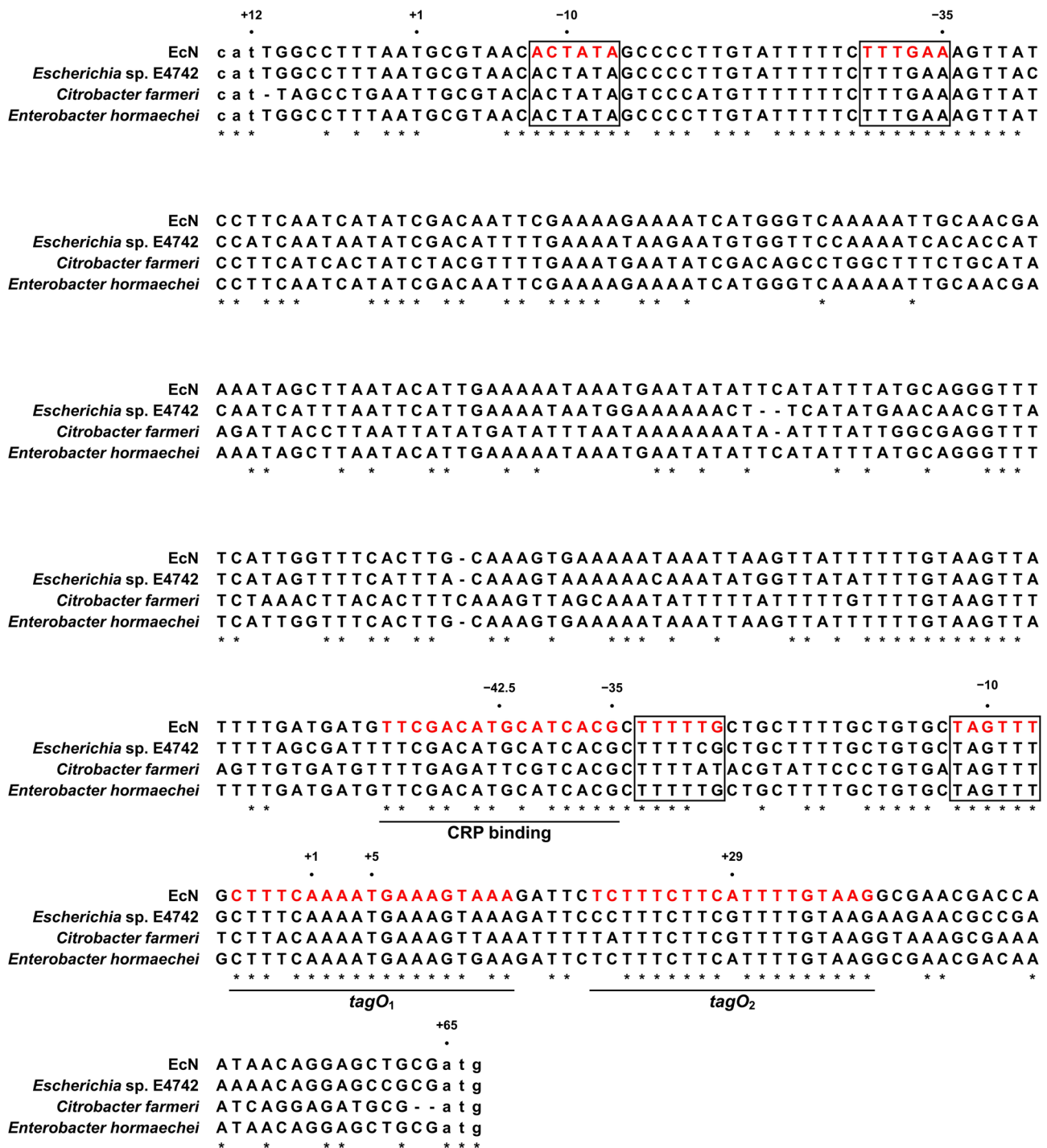
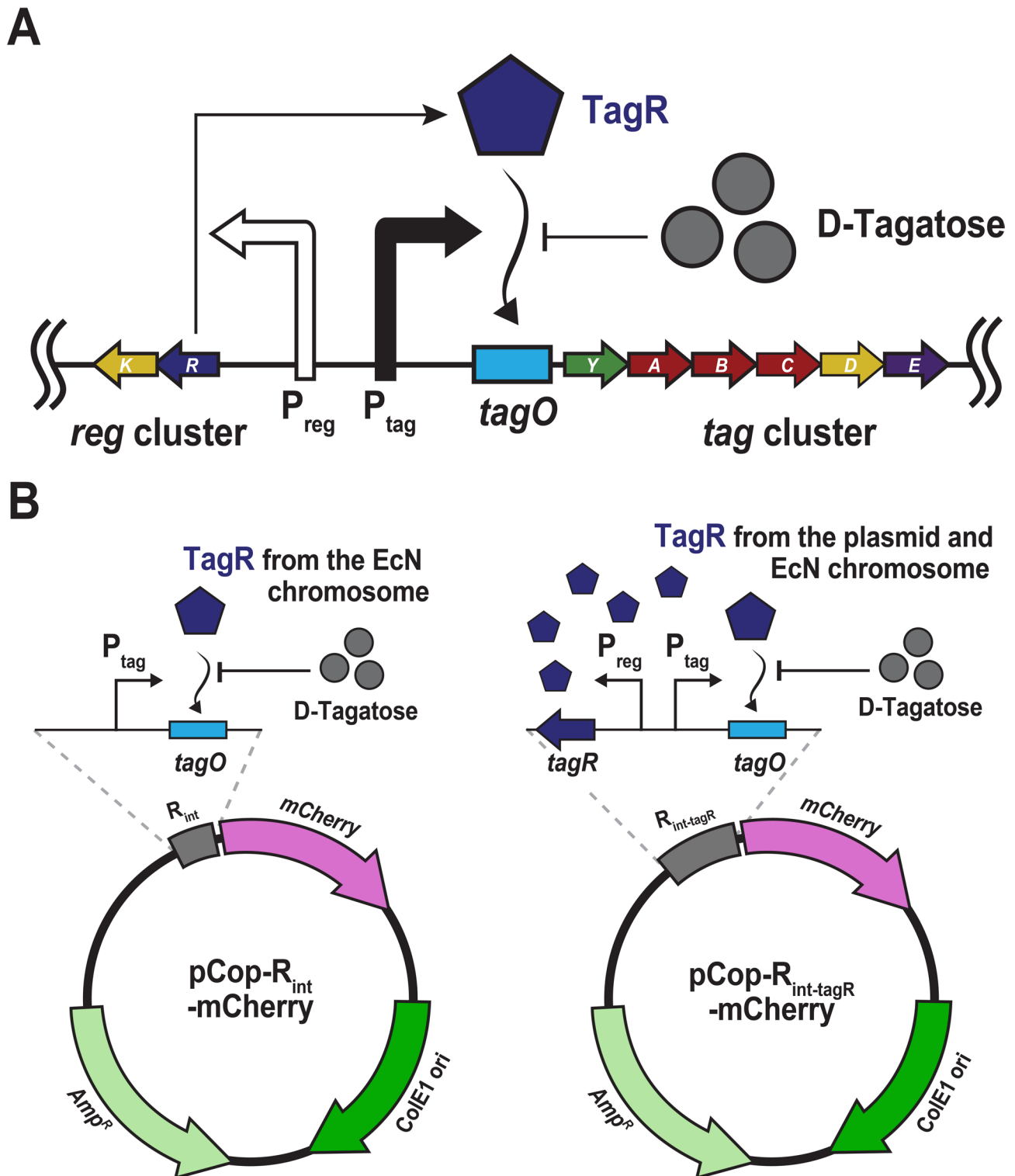


Fig. 2 Multiple sequence alignment of the *tagR*–*tagY* intergenic region. The initiation codons for *tagR* and *tagY* are located at the beginning and end of the sequence, respectively. Their TSSs are positioned as +1 and marked with arrows indicating transcription direction. The predicted –10 and –35 regions of the promoters (P_{tag} and P_{reg}) are boxed. The predicted CRP binding site is underlined, whereas the operator sites (*tagO*₁ and *tagO*₂) recognized by TagR are indicated by horizontal arrows indicating their directionality. Asterisks denote conserved nucleotides among the aligned sequences. Conserved sequences from the multiple alignment at promoters and transcription factor binding sites are highlighted in red. Nucleotide sequences were obtained from the genome sequences of *E. coli* Nissle 1917 (EcN) (RefSeq accession NZ_CP058217), *Escherichia* sp. E4742 (GenBank accession CP040443), *Citrobacter farmeri* (CP022695), and *Enterobacter hormaechei* (CP056649)



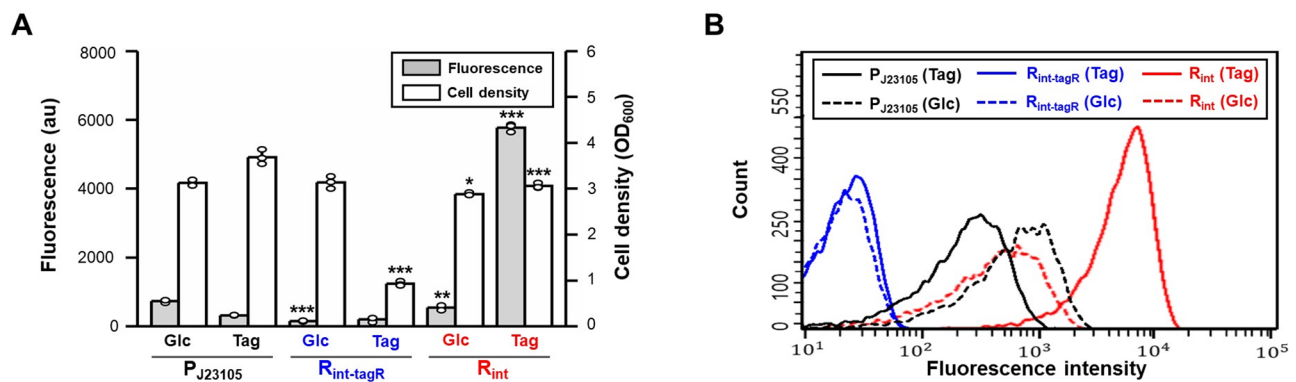


Fig. 4 Effects of *tagR* overexpression on mCherry expression. **(A)** Quantification of mCherry fluorescence and cell density in EcN cells harboring pCop-P_{J23105}-mCherry, pCop-R_{int}-mCherry, or pCop-R_{int-tagR}-mCherry. Individual data points are shown as open circles. Pairwise comparisons were performed between the P_{J23105} construct and each of the other constructs under the same carbon source. The *p*-values were adjusted for multiple comparisons using Dunnett's test (**p* < 0.05, ***p* < 0.01, ****p* < 0.001). **(B)** Single-cell analysis of mCherry expression by flow cytometry. Cells were grown in M9 minimal medium supplemented with either 20 mM glucose (Glc) or 20 mM D-tagatose (Tag) as the sole carbon source for 48 h. Error bars denote the standard deviation of the mean from three independent cultivations

Moreover, cells expressing pCop-R_{int}-mCherry showed a distinct and homogeneous increase in fluorescence intensity in the presence of D-tagatose (Fig. 4B), indicating that D-tagatose effectively acted as a uniform inducer, leading to consistent activation of mCherry expression across the cell population.

Performance of the tagatose-inducible expression system

Kinetic analysis of mCherry expression in EcN cells expressing the pCop-R_{int}-mCherry construct was performed to investigate the dose-dependent relationship between D-tagatose availability and gene expression. Transformed EcN cells were grown in minimal medium with varying concentrations of D-tagatose (4–20 mM) as the sole carbon source. This range was chosen based on our previous growth assay [17], which demonstrated that wild-type EcN showed reduced growth and incomplete D-tagatose utilization when cultured in minimal medium containing D-tagatose at concentrations above 40 mM, possibly due to metabolic saturation or inhibitory effects. To avoid these issues and ensure a dynamic, dose-responsive induction range, we limited the D-tagatose concentration to 0–20 mM. Increasing D-tagatose concentrations resulted in both higher fluorescence levels and cell growth (Fig. 5A), indicating that the expression system was robust and did not impose a considerable metabolic burden on the cells. After 48 h of cultivation, mCherry expression levels exhibited a strong positive correlation with D-tagatose concentration, with a correlation coefficient (*r*) of 0.95 (Fig. 5B).

Flow cytometry analysis (Fig. 5B) provided further insights into the homogeneity of gene expression. As D-tagatose concentrations increased, the fluorescence intensity distribution curves became narrow and tall,

indicating that high inducer levels not only enhanced the mean expression levels but also improved the synchrony of gene expression across the cell population. These results demonstrate that the developed tagatose-inducible expression system enables dose-dependent control of gene expression in EcN under the conditions tested.

Catabolite repression-enabled auto-induction for protein expression

To enhance cell growth while achieving target protein expression, we adapted the tagatose-inducible system to auto-induction media containing both glucose and D-tagatose. The location of the CRP-binding site near the P_{tag} promoter (Fig. 1B) suggests that the *tag* cluster would be subject to catabolite repression. Based on this, we hypothesized that, in the presence of glucose, EcN would grow rapidly with minimal target production. Upon glucose depletion, D-tagatose would be utilized by cells to trigger target protein production.

To test this hypothesis, we performed a kinetic analysis of mCherry expression and growth in EcN cells harboring the pCop-R_{int}-mCherry construct. Cells were cultured in minimal medium supplemented with 20 mM glucose and varying concentrations of D-tagatose (0–20 mM). In media containing only glucose, mCherry expression increased after glucose was nearly depleted (Fig. 6A). In mixed glucose and D-tagatose media, glucose was rapidly consumed within approximately 8 h, followed by gradual consumption of D-tagatose. During the glucose consumption phase, cell density increased rapidly, whereas mCherry expression remained low. Once glucose was depleted and D-tagatose consumption started, mCherry expression began to increase, although cell density showed little further

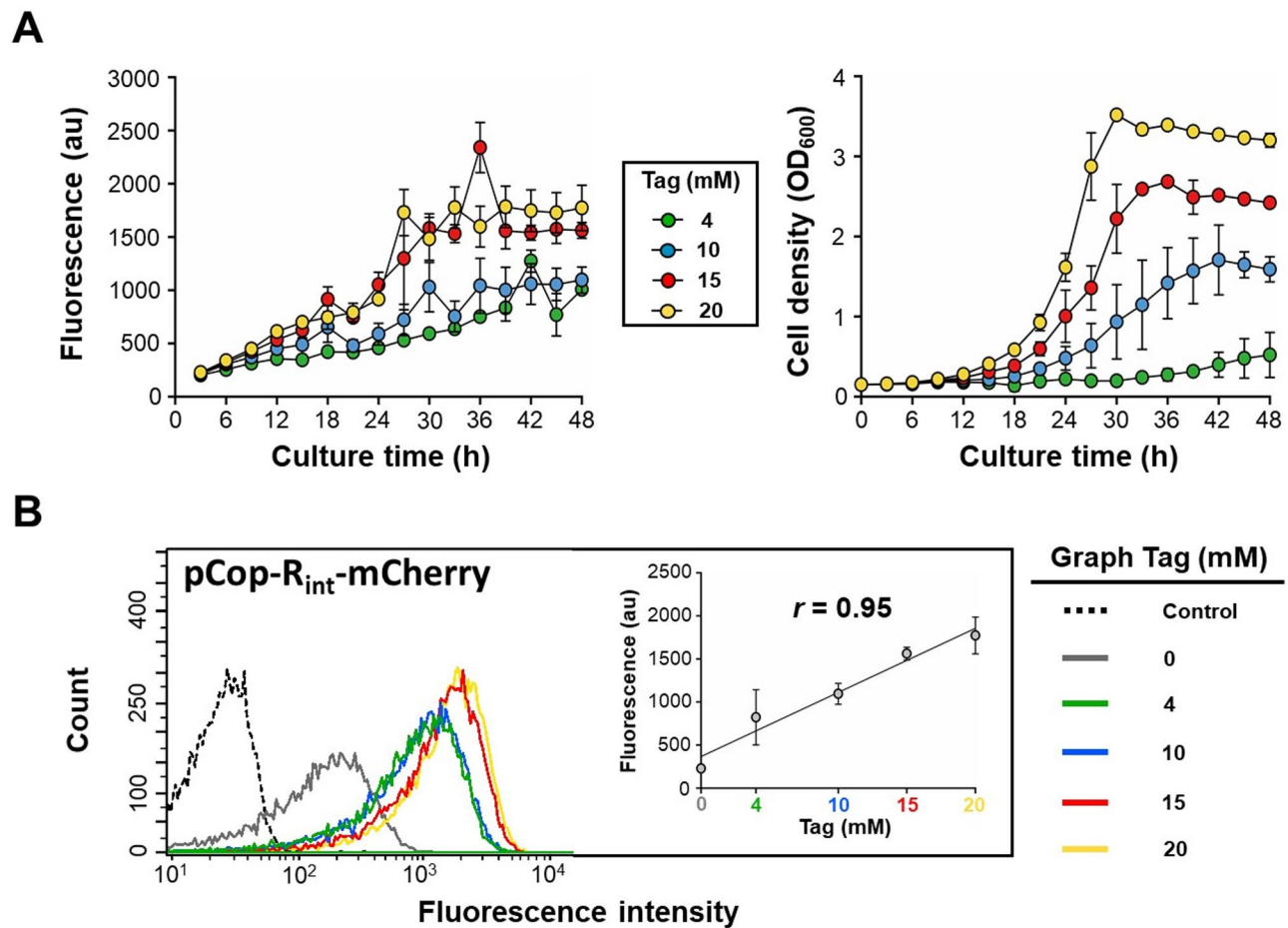


Fig. 5 Tagatose dose-dependent mCherry expression from the tagatose-inducible expression plasmid in media containing D-tagatose. **(A)** Time-course analysis of mCherry fluorescence intensity (left) and cell density (right) in EcN cells harboring the pCop-R_{int}-mCherry plasmid. Cells were cultured in M9 minimal media containing varying concentrations of D-tagatose (Tag). Error bars denote the standard deviation of the mean from three independent cultivations. **(B)** Single-cell analysis of mCherry expression by flow cytometry. The fluorescence intensity distribution shows the population of EcN cells harboring pCop-R_{int}-mCherry at the end of cultivation (48 h). The transformed cells grown with 20 mM glucose (shown in gray) serve as the 0 mM D-tagatose condition. The inset bar chart shows the mean fluorescence intensity values $\pm$ SD for each condition. Wild-type EcN cells grown with 20 mM glucose (dotted line) were used as a negative control

increase. This suggests that catabolite repression is reversible and the *tag* cluster can be derepressed when glucose is exhausted from the medium. At the end of the cultivation period, mCherry expression was strongly correlated with D-tagatose concentration ($r=0.92$) (Fig. 6B), demonstrating that mCherry expression is highly dependent on D-tagatose levels after glucose depletion. These results demonstrated that the *tag* cluster was subject to catabolite repression and validated the suitability of the tagatose-inducible system for the auto-induction strategy. This method effectively separated the growth and expression phases, allowing rapid cell growth during glucose consumption and robust target protein expression during D-tagatose utilization.

Overexpression of Recombinant proteins

The tagatose-inducible expression system was used to overexpress two recombinant proteins, MBP-hEGF and PETase, in auto-induction media. hEGF is a pharmaceutical protein that promotes the proliferation of epidermal cells [37] and is widely used in medicine and cosmetics. Previously, EcN was engineered to secrete hEGF using an IPTG-inducible promoter [38, 39]. PETase, an industrial enzyme derived from *Ideonella sakaiensis*, degrades PET plastics [40] and has potential applications in PET recycling and bioremediation.

For MBP-hEGF expression, the pCop-R_{int}-MBP-hEGF plasmid was constructed containing the MBP and hEGF sequences downstream of the R_{int} promoter. Similarly, the pCop-R_{int}-PETase plasmid was constructed to express PETase. Each plasmid was introduced into EcN cells, which were aerobically grown

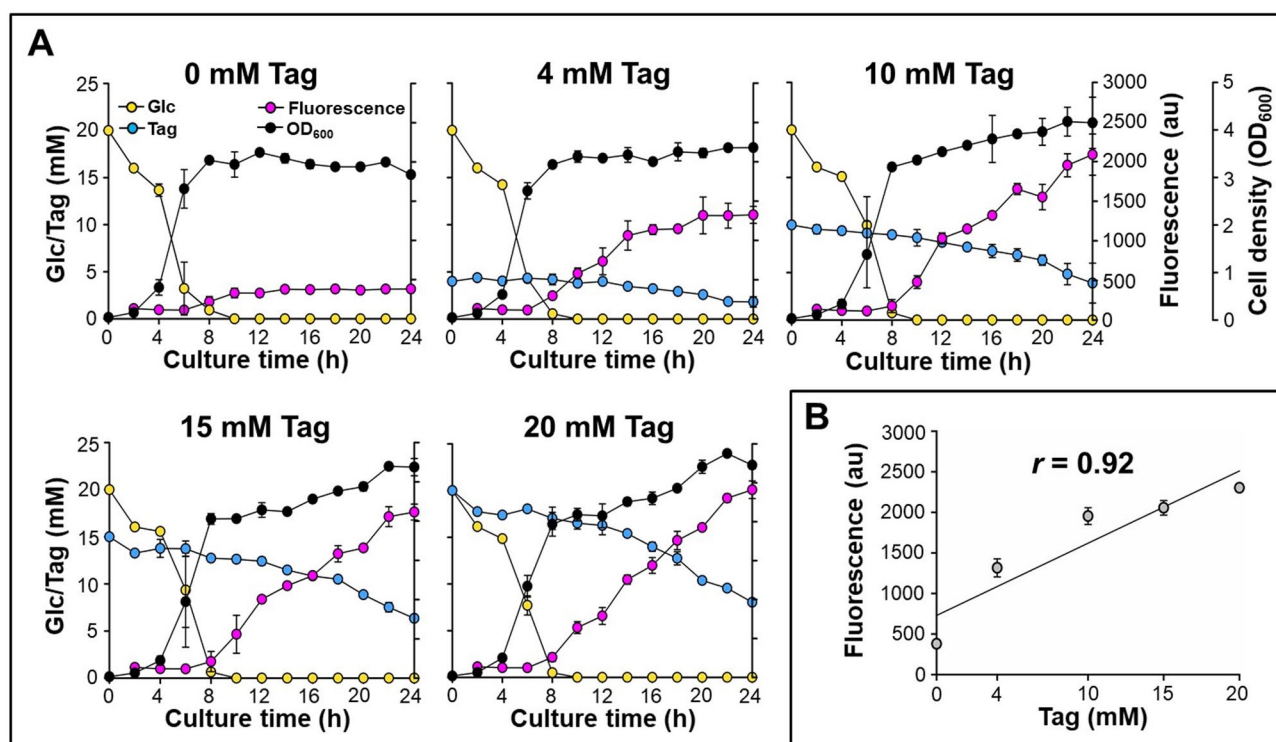


Fig. 6 Tagatose dose-dependent mCherry expression from the tagatose-inducible expression plasmid in media containing both D-tagatose and glucose. **(A)** Time-course analysis of mCherry fluorescence, cell density, and residual concentrations of glucose and D-tagatose in EcN cells harboring the pCop-R_{int}-mCherry plasmid. Cells were cultured in M9 minimal media containing 20 mM glucose (Glc) and varying concentrations of D-tagatose (Tag). **(B)** Correlation between D-tagatose concentration and mCherry fluorescence at the end of the cultivation (24 h). Error bars denote the standard deviation of the mean from two independent cultivations

in minimal media containing a combination of glucose and varying concentrations of D-tagatose for 48 h. Western blot analysis revealed a strong correlation between MBP-hEGF expression levels and the D-tagatose concentrations tested (0–20 mM) ($r = 0.93$) (Fig. 7A). PETase production also showed a strong correlation with D-tagatose concentration (0–4 mM) ($r = 0.93$) (Fig. 7B), although it saturated beyond 4 mM.

To evaluate the performance of the tagatose-inducible expression system under anaerobic conditions, EcN cells harboring the pCop-R_{int}-MBP-hEGF plasmid were cultivated anaerobically in minimal medium containing a combination of glucose and varying concentrations of D-tagatose. As expected, cell growth was the highest when 20 mM glucose was used as the sole carbon source, followed by the combination of 4 mM glucose and 20 mM D-tagatose, and was the lowest when 20 mM D-tagatose was used alone (Fig. S2A). Notably, MBP-hEGF expression was undetectable in glucose-only cultures but was considerably induced in cultures containing either D-tagatose alone or a combination of glucose and D-tagatose (Fig. S2B). Under anaerobic conditions, MBP-hEGF expression showed a strong positive correlation with D-tagatose concentration (0–20 mM) in the presence of 20 mM

glucose ($r = 0.91$) (Fig. 7C). These examples illustrate that the tagatoseinducible system supports recombinant protein production under both aerobic and anaerobic conditions, suggesting its utility for applications that require controlled expression in different environments.

Conclusion

EcN, a well-established probiotic, has emerged as a versatile microbial chassis for use in living therapeutics, drug delivery, and industrial bioproduction. However, its clinical and industrial applications are limited by the lack of reliable inducible expression systems. In this study, we developed a novel tagatose-inducible expression system for EcN, enabling dose-dependent control with uniform gene expression across the cell population. Given that D-tagatose is a metabolizable sugar with established safety in humans, this system holds potential for in vivo applications such as therapeutic protein delivery and biosensing within the human gut microbiome, although further in vivo validation will be necessary to confirm its clinical utility. The tagatoseinducible system provides an additional tool to support synthetic biology and metabolic engineering efforts in EcN. Furthermore, the identification

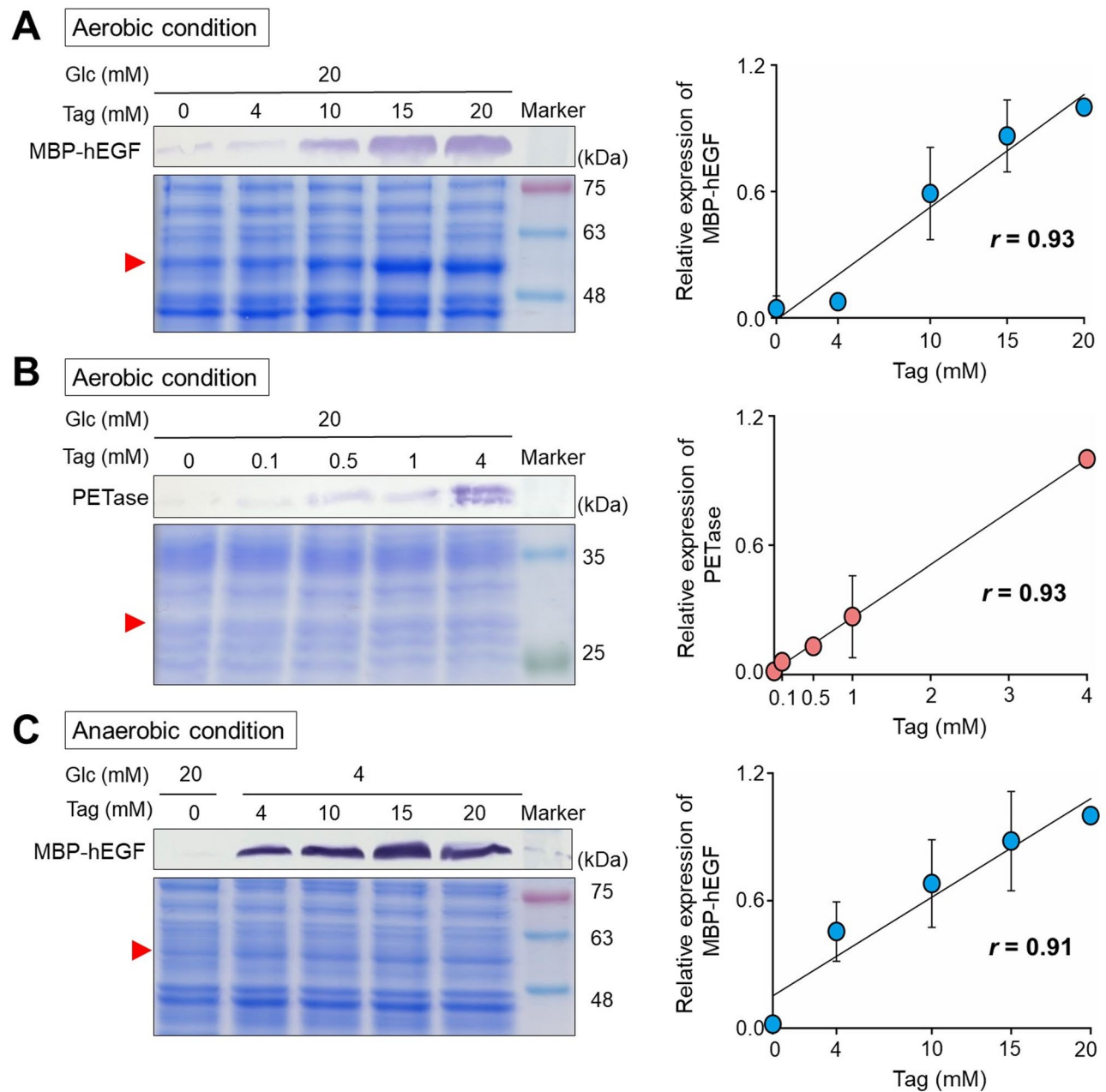


Fig. 7 Tagatose dose-dependent expression of recombinant proteins from the tagatose-inducible expression plasmid. **(A)** Expression of MBP-hEGF fusion protein in EcN harboring pCop- R_{int} -MBP-hEGF under aerobic conditions. **(B)** Expression of PETase in EcN harboring pCop- R_{int} -PETase under aerobic conditions. **(C)** Expression of MBP-hEGF fusion protein in EcN harboring pCop- R_{int} -MBP-hEGF under anaerobic conditions. The transformed EcN cells were cultured in media containing glucose (Glc) and varying concentrations of D-tagatose (Tag) for 48 h. The gel images (left) show the western blot analysis of the target protein (upper section) and SDS-PAGE analysis of total protein extracts (lower section), with the red arrowhead indicating the target protein. The graphs (right) display the relative expression of the target protein quantified from western blot analysis. Error bars denote the standard deviation of the mean from independent cultivations

of key regulatory elements within the tagatose catabolic gene cluster provides a foundation for optimizing gene expression and broadening its applicability in precision medicine and industrial bioproduction.

It should be noted that the functionality of the D-tagatose-inducible system is inherently dependent on the presence of the native tagatose catabolic gene

cluster, which enables both the uptake and degradation of D-tagatose. As previously reported [17], the *tag* cluster is prevalent in *E. coli* B2 phylogroup strains, including EcN, but is absent in non-B2 strains such as *E. coli* K-12. Therefore, introducing the developed plasmid alone into K-12 strains is insufficient to activate the P_{tag} promoter. To extend the applicability of

this system to other *E. coli* backgrounds, the relevant tagatose metabolism genes need to be co-introduced. Follow-up studies are required to develop a version of the system that operate across diverse *E. coli* strains.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12934-025-02771-4>.

Supplementary material 1: Tables S1 & S2 and Figs. S1 & S2

Supplementary material 2: pCOP-X-mCherry.gb

Acknowledgements

Not applicable.

Author contributions

SHY and SMY conceived and supervised the project. KP performed the genetic manipulation, aerobic cultivations, and protein expression analyses. YK performed the dRNA-seq analysis and anaerobic cultivations. DK analyzed the regulatory elements. SMO performed flow cytometry analyses. TJK contributed to aerobic cultivations. KP, YK, and SHY wrote the manuscript. All authors reviewed and approved the final manuscript.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Ministry of Science and ICT (MSIT) [RS-2024-00401518 to SHY and NRF-2022R1A2C2004292 to SMY] and by the Strategic Initiative for Microbiomes in Agriculture and Food funded by the Ministry of Agriculture, Food, and Rural Affairs [RS-2018-IP918012 to SHY].

Data availability

The dRNA data were deposited in the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) database under the accession number GSE280463. All data generated or analysed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 13 March 2025 / Accepted: 16 June 2025

Published online: 04 July 2025

References

1. Dosoky NS, May-Zhang LS, Davies SS. Engineering the gut microbiota to treat chronic diseases. *Appl Microbiol Biotechnol*. 2020;104:7657–71.
2. Aggarwal N, Breedon AME, Davis CM, Hwang IY, Chang MW. Engineering probiotics for therapeutic applications: recent examples and translational outlook. *Curr Opin Biotechnol*. 2020;65:171–9.
3. Jacobi CA, Malfertheiner P. *Escherichia coli* Nissle 1917 (Mutaflor): new insights into an old probiotic bacterium. *Dig Dis*. 2011;29:60–7.
4. Behnsen J, Deriu E, Sassone-Corsi M, Raffatellu M. Probiotics: properties, examples, and specific applications. *Cold Spring Harb Perspect Med*. 2013;3:a010074.
5. Yu M, Hu S, Tang B, Yang H, Sun D. Engineering *Escherichia coli* Nissle 1917 as a microbial chassis for therapeutic and industrial applications. *Biotechnol Adv*. 2023;67:108202.
6. Lynch JP, Goers L, Lesser CF. Emerging strategies for engineering *Escherichia coli* Nissle 1917-based therapeutics. *Trends Pharmacol Sci*. 2022;43:772–86.
7. Fiege K, Frankenberg-Dinkel N. Construction of a new T7 promoter compatible *Escherichia coli* Nissle 1917 strain for recombinant production of heme-dependent proteins. *Microb Cell Fact*. 2020;19:190.
8. Li X, Yu Y, Tang J, Gong B, Li W, Chen T, Zhou X. The construction of a dual-functional strain that produces both polysaccharides and sulfotransferases. *Biotechnol Lett*. 2021;43:1831–44.
9. Effendi SSW, Ng IS. Reprogramming T7RNA polymerase in *Escherichia coli* Nissle 1917 under specific Lac Operon for efficient p-coumaric acid production. *ACS Synth Biol*. 2022;11:3471–81.
10. Tan SI, Ng IS. New insight into plasmid-driven T7 RNA polymerase in *Escherichia coli* and use as a genetic amplifier for a biosensor. *ACS Synth Biol*. 2020;9:613–22.
11. Zhang ZX, Nong FT, Wang YZ, Yan CX, Gu Y, Song P, Sun XM. Strategies for efficient production of recombinant proteins in *Escherichia coli*: alleviating the host burden and enhancing protein activity. *Microb Cell Fact*. 2022;21:191.
12. Cui M, Pang G, Zhang T, Sun T, Zhang L, Kang R, Xue X, Pan H, Yang C, Zhang X, et al. Optotheranostic nanosystem with phone visual diagnosis and optogenetic microbial therapy for ulcerative colitis at-home care. *ACS Nano*. 2021;15:7040–52.
13. Pan H, Li L, Pang G, Han C, Liu B, Zhang Y, Shen Y, Sun T, Liu J, Chang J, Wang H. Engineered NIR light-responsive bacteria as anti-tumor agent for targeted and precise cancer therapy. *Chem Eng J*. 2021;426:130842.
14. Oh DK. Tagatose: properties, applications, and biotechnological processes. *Appl Microbiol Biotechnol*. 2007;76:1–8.
15. Levin GV. Tagatose, the new GRAS sweetener and health product. *J Med Food*. 2002;5:23–36.
16. Hans Bertelsen HAMT. Fermentation of D-tagatose by human intestinal bacteria and dairy lactic acid bacteria. *Microb Ecol Health Dis*. 2001;13:87–95.
17. Ha J, Kim D, Yeom J, Kim Y, Yoo SM, Yoon SH. Identification of a gene cluster for D-tagatose utilization in *Escherichia coli* B2 phylogroup. *iScience*. 2022;25:105655.
18. Jung YJ, Park KH, Jang TY, Yoo SM. Gene expression regulation by modulating Hfq expression in coordination with tailor-made sRNA-based knockdown in *Escherichia coli*. *J Biotechnol*. 2024;388:1–10.
19. Kim B-L, Lee EBJ, Kim H-W, Ahn CS, Lee JO, Jung HW, Lee J-K, Kim EG. Study on the soluble expression of recombinant human epidermal growth factor using various fusion partners in *Escherichia coli*. *Korean J Biotechnol Bioeng*. 2008;23:205–12.
20. Schneider CA, Rasband WS, Eliceiri KW. NIH image to imagej: 25 years of image analysis. *Nat Methods*. 2012;9:671–5.
21. Kim S, Jeong H, Kim EY, Kim JF, Lee SY, Yoon SH. Genomic and transcriptomic landscape of *Escherichia coli* BL21(DE3). *Nucleic Acids Res*. 2017;45:5285–93.
22. Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R, et al. The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature*. 2010;464:250–5.
23. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
24. Förstner KU, Vogel J, Sharma CM. READemption-a tool for the computational analysis of deep-sequencing-based transcriptome data. *Bioinformatics*. 2014;30:3421–3.
25. Yu SH, Vogel J, Förstner KU. ANNOgesic: a Swiss army knife for the RNA-seq based annotation of bacterial/archaeal genomes. *Gigascience*. 2018;7:gij096.
26. Freese NH, Norris DC, Loraine AE. Integrated genome browser: visual analytics platform for genomics. *Bioinformatics*. 2016;32:2089–95.
27. Liu B, Yang F, Huang DS, Chou KC. iPromoter-2L: a two-layer predictor for identifying promoters and their types by multi-window-based PseKNC. *Bioinformatics*. 2018;34:33–40.
28. Solovyev V, Salamov A. Automatic annotation of microbial genomes and metagenomic sequences. In *Metagenomics and its applications in agriculture, biomedicine and environmental studies*. Edited by Li RW. Hauppauge, NY: Nova Science Publishers Inc; 2011: 61–78.
29. Chen Y, Lin YC, Luo Y, Cai X, Qiu P, Cui S, Wang Z, Huang HY, Huang HD. Quantitative model for genome-wide Cyclic AMP receptor protein binding site identification and characteristic analysis. *Brief Bioinform*. 2023;24:bbad138.
30. Tsai MJ, Wang JR, Yang CD, Kao KC, Huang WL, Huang HY, Tseng CP, Huang HD, Ho SY. PredCRP: predicting and analysing the regulatory roles of CRP from its binding sites in *Escherichia coli*. *Sci Rep*. 2018;8:951.
31. Lee W, Park B, Han K. Sequence-based prediction of putative transcription factor binding sites in DNA sequences of any length. *IEEE/ACM Trans Comput Biol Bioinform*. 2017;15:1461–9.

32. Madeira F, Madhusoodanan N, Lee J, Eusebi A, Niewielska A, Tivey ARN, Lopez R, Butcher S. The EMBL-EBI job dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res.* 2024;52:W521–5.
33. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* 2018;27:135–45.
34. Demerec M, Adelberg EA, Clark AJ, Hartman PE. A proposal for a uniform nomenclature in bacterial genetics. *Genetics.* 1966;54:61–76.
35. Ulanova D, Kitani S, Fukusaki E, Nihira T. SdrA, a new deor family regulator involved in *Streptomyces avermitilis* morphological development and antibiotic production. *Appl Environ Microbiol.* 2013;79:7916–21.
36. Ray WK, Larson TJ. Application of agar repressor and dominant repressor variants for verification of a gene cluster involved in N-acetylgalactosamine metabolism in *Escherichia coli* K-12. *Mol Microbiol.* 2004;51:813–26.
37. Carpenter G, Cohen S. Epidermal growth factor. *J Biol Chem.* 1990;265:7709–12.
38. Choi HJ, Ahn JH, Park SH, Do KH, Kim J, Moon Y. Enhanced wound healing by recombinant *Escherichia coli* Nissle 1917 via human epidermal growth factor receptor in human intestinal epithelial cells: therapeutic implication using Recombinant probiotics. *Infect Immun.* 2012;80:1079–87.
39. Yu M, Kim J, Ahn JH, Moon Y. Nononcogenic restoration of the intestinal barrier by *E. coli*-delivered human EGF. *JCI Insight.* 2019;4:e125166.
40. Yoshida S, Hiraga K, Takehana T, Taniguchi I, Yamaji H, Maeda Y, Toyohara K, Miyamoto K, Kimura Y, Oda K. A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science.* 2016;351:1196–9.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.