Nucleic Acids Research, 2024, **52**, 11552–11570
https://doi.org/10.1093/nar/gkae773
Advance access publication date: 11 September 2024
Gene regulation, Chromatin and Epigenetics

A single rare σ^{70} variant establishes a unique gene expression pattern in the *E. coli* pathobiont LF82

Melissa Arroyo-Mendoza[1,2], Alexandra Proctor[2], Abraham Correa-Medina[1], Sarah DeWolf[2], Meghan Wymore Brand[2], Virginia Rosas[1], Hernan Lorenzi[3], Michael J. Wannemuehler[2], Gregory J. Phillips[2,*] and Deborah M. Hinton[1,*]

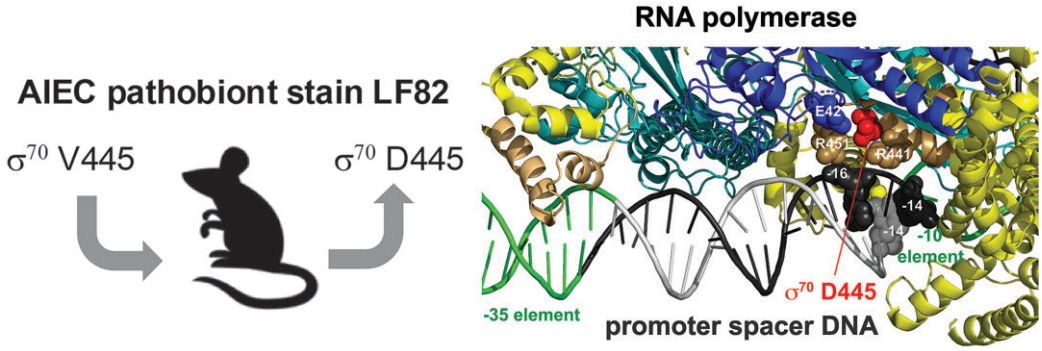
[1]Gene Expression and Regulation Section, Laboratory of Biochemistry and Genetics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, 8 Center Dr., Bethesda, MD, USA
[2]Department of Veterinary Microbiology and Preventative Medicine, College of Veterinary Medicine, Iowa State University, Ames, IA, USA
[3]TriLab Bioinformatics Group, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, 8 Center Dr., Bethesda, MD, USA

[*]To whom correspondence should be addressed. Tel: +1 301 496 9885; Email: dhinton@helix.nih.gov
Correspondence may also be addressed to Gregory J. Phillips. Tel: +1 706 542 3438; Email: greg.phillips@uga.edu
Present addresses:
Alexandra Proctor, South First Street, Long Grove, IA, USA.
Abraham Correa-Medina, University of Washington School of Medicine, University of Washington, Seattle, WA, USA.
Meghan Wymore Brand, Virus and Prion Research Unit, National Animal Disease Center, USDA-ARS, Ames, IA, USA.
Gregory J. Phillips, Department of Infectious Diseases, University of Georgia, Athens, GA, USA.

Abstract

LF82, an adherent-invasive *Escherichia coli* (AIEC) pathobiont, is associated with Crohn's disease, an inflammatory bowel disease of unknown etiology. Although AIEC phenotypes differ from those of 'commensal' or pathogenic *E. coli*, work has failed to identify genetic features accounting for these differences. We have investigated a natural, but rare, single nucleotide polymorphism (SNP) in LF82 present within the highly conserved *rpoD* gene, encoding σ^{70} [primary sigma factor, RNA polymerase (RNAP)]. We demonstrate that σ^{70} D445V results in transcriptomic and phenotypic changes consistent with LF82 phenotypes, including increased antibiotic resistance and biofilm formation and increased capacity for methionine biosynthesis. RNA-seq analyses comparing σ^{70} V445 versus σ^{70} D445 identified 24 genes upregulated by σ^{70} V445 in both LF82 and the laboratory *E. coli* K-12 strain MG1655. Using *in vitro* transcription, we demonstrate that σ^{70} D445V directly increases transcription from promoters for several of the up-regulated genes and that the presence of a 16 bp spacer and -14 G:C is associated with this increase. The position of D445V within RNAP suggests that it could affect RNAP/spacer interaction. Our work represents the first identification of a distinguishing SNP for this pathobiont and suggests an underrecognized mechanism by which pathobionts and strain variants can emerge.

Graphical abstract



Introduction

Inflammatory bowel diseases (IBD), such as Crohn's disease (CD) and ulcerative colitis (UC), are chronic, relapsing, and immunological-mediated disorders consisting of prolonged bouts of inflammation of the intestinal mucosa. In addition to negatively impacting quality of life, CD patients have an elevated risk of developing colorectal and small bowel cancers as well as complications from surgery. As multiple factors are implicated in the etiology of CD, including diet, host genetic susceptibility, and environmental and immunologic factors, new insights are needed to better understand the cause of the disease and to develop a cure (1). Several studies have identified associations of disease with the presence of adherent-invasive *E. coli* (AIEC) pathobionts, which have been reported to be

Received: June 12, 2024. **Revised:** August 8, 2024. **Editorial Decision:** August 10, 2024. **Accepted:** August 28, 2024
Published by Oxford University Press on behalf of Nucleic Acids Research 2024.

elevated by 1–2 orders of magnitude in CD patients where they are thought to be crucial to the inflammatory processes present in this disease (2,3).

LF82 is the prototypical AIEC strain and has received considerable attention in efforts to better understand the role of AIEC in CD (3,4). This pathobiont is a member of the B2 phylogenetic group and possesses unique phenotypes that distinguish it from most other *Escherichia coli* strains (3,4). While LF82 does not express known toxins or virulence factors found in frank *E. coli* pathogens, it can attach to and invade intestinal epithelial cells, survive in macrophages, and induce secretion of tumor necrosis factor-alpha (TNF-α) and other pro-inflammatory cytokines (5–7). LF82 binds to CECAM6 (carcinoembryonic antigen–related cell adhesion molecule 6), which is expressed on epithelial cells, especially in inflamed intestinal tissue (8). Like other AIEC strains, LF82 can also oxidize a variety of substrates to outcompete other members of the gut microbiota (9). Despite these associations with CD, it remains unclear how LF82 emerged as a gut-associated pathobiont and which features of its genetic makeup contribute to its unique phenotypes. Specific genetic signatures that differentiate AIEC strains from other *E. coli* isolates have also not been identified (10–13), suggesting that AIEC strains have emerged independently multiple times (14).

To better understand the processes that contribute to the emergence of AIEC strains, a murine host-to-host transmission model was previously used to identify genetic changes associated with adaptation to the gut across multiple hosts (14). This effort identified AIEC variants with altered motility and acetate metabolism that improved transmission. In contrast to identifying genetic changes associated with host-to-host transmission, we sought to use a defined microbiota mouse model to establish chronic, long-term colonization of LF82 without requiring re-colonization into new hosts. For this, we introduced LF82 into altered Schaedler flora (ASF) mice, which represent a gnotobiotic animal model where the animals are colonized with only 8 bacterial species, excluding members of the *Proteobactericeae* (15,16). In contrast to conventional mice with complex microbiota, ASF mice maintain LF82 at consistent levels without the need for continual reacquisition or disruptive antibiotic treatment (17). Following long term colonization, we employed whole genome sequencing to identify single nucleotide polymorphisms (SNPs) in LF82 that occurred as the AIEC pathobiont adapted to the ASF murine gut. Here we focus on an extremely rare SNP within RNA polymerase (RNAP) that is present in LF82.

Eubacterial RNAP is composed of a core of five subunits (α_2, β, β', ω) that is responsible for the bulk of mRNA and structural RNA synthesis within the cell. However, to initiate transcription core RNAP requires a σ factor to form the holoenzyme, which recognizes and binds to specific promoter sequences (18). The primary σ factor (σ^{70} in *E. coli*) is the essential 'housekeeping' transcription factor that is necessary for bacterial growth. The major promoter binding sites for σ^{70} are the -35 element, the extended −10 element, and the −10 element, which are typically located from −30 to −35, from −15 to −14, and from −12 to −7, respectively, relative to the +1 transcription start site (TSS). Additionally, the spacing distance between the −10 and −35 promoter motifs is important, with an acceptable spacer length of 16–18 bp. Interactions of RNAP with the −35 and −10 elements has been extensively investigated. However, how the sequence and length of the spacer affect promoter activity has been a less-explored area of research.

LF82 encodes a variant of σ^{70} in which residue V445 differs from the highly conserved D445 found in thousands of σ^{70} sequences throughout *Enterobacteriales*. Strikingly, we show that this unique σ^{70} variant 'reverts' back to the highly conserved D445 after passage through ASF mice. Presumably this represents the conversion of LF82 to a more 'commensal-like' existence within the gut of the ASF mouse.

To better understand the significance of this variation in a highly conserved housekeeping gene, we constructed the D445V mutant within a nonpathogenic K-12 strain of *E. coli* (MG1655) as well as the V445D substitution within LF82 and performed RNA-seq analyses along with functional assays to determine how this single substitution affects gene expression and strain phenotypes. Our results are consistent with the idea that σ^{70} V445 contributes to the lifestyle of LF82, including increases in biofilm and antibiotic resistance. To understand whether this substitution affects transcription directly, we purified σ^{70} V445 (the rare LF82 variant) and compared its activity to σ^{70} D445 (the common version in *E. coli*) in *in vitro* transcription assays. Based on RNAP structures (19), the position of σ^{70} D445 posits a possible effect of RNAP interaction with promoter spacer DNA. Using multiple promoters, we show that a promoter spacer of 16 bp and a G:C bp at position -14 correlates with increased transcription *in vitro* from several promoters upstream of genes whose expression increases in the RNA-seq analyses. Our results suggest that the σ^{70} D445V substitution affects the RNAP–spacer interactions, resulting in a direct increase in the expression of specific genes by altering transcription initiation. Importantly, the work identifies the first distinguishing SNP for LF82.

More broadly, our results further suggest that mutations within the highly conserved *rpoD* housekeeping gene may represent an underexplored mechanism contributing to bacterial adaptation to new environments and the emergence of new pathotypes.

Materials and methods

Bacterial strains

E. coli LF82, originally isolated from a chronic ileal lesion of a patient with CD (20), is maintained in the Wannemuehler lab. BL21(DE3)/pLysE (21) was used for transformation of the pETσ^{70}*rpoD* D445V plasmid.

The *E. coli* str. K-12 substr. MG1655 [F- lambda- *ilvG-rfb*-50 *rph*-1] was a gift from Michael Cashel (NIH). MG1655^{D445V} (also designated as NB211) was constructed using a scarless genome editing approach that has been described previously (22). Briefly, this included cloning a 1 kbp synthetic DNA fragment gBlock (Integrated DNA Technologies, Coralville, IA), including nucleotide position 1334 of *rpoD* from MG1655 where GTT was changed to GAT, flanked by ∼500 bp of homology. The gBlock was used as a template for a PCR reaction using primers designed to introduce Nt.BspQI restriction enzyme recognition sites into the ends of the amplification products (*rpoD*-BspQI.S: 5′-TCGACATTCCAAGGGGAAGAGCGCGTTGAGCAGTGCAAA-3′ and *rpoD*-BspQI.AS: 5′-ACTCATCGACTAGGTGAAGAGCCGATCCGGCCTACCGATTA-3′.
The PCR products were digested with Nt.BspQI

[New England Biolabs (NEB), Ipswich, MA)], which facilitated cloning into an R6K-based suicide vector constructed specifically for allelic exchange (pRIPR). pRIPR imparts resistance to chloramphenicol (Cam^R), includes the 18-bp recognition site for the I-SceI nuclease (5′-TAGGGATAACAGGGTAAT-3'), and allows insertion of DNA at Nt.BspQI sites. The recombinant pRIPR plasmid transformed into the diaminopimelic acid (DAP) auxotroph donor strain MFD*pir* (23), which was used to introduce the R6K plasmid into MG1655 by conjugation. Cam^R colonies that grew in the absence of DAP were selected at 37°C, yielding recombinants where the suicide vector had integrated into the *E. coli* chromosome at *rpoD*. These colonies were subsequently transformed with the helper plasmid pSceH, a derivative of a vector that expresses the I-SceI nuclease under the control of the *tetR* promoter from a temperature-sensitive pSC101-derivative plasmid (24). Ampicillin-resistant (Amp^R)transformants were selected at 30°C in the presence of anhydrotetracycline to induce synthesis of I-SceI. The resulting colonies were then screened to identify recombinants that had become Cam-sensitive (Cam^S). Multiple Cam^S recombinants were tested by Sanger sequencing (Iowa State University's DNA Facility) for a ∼600-bp PCR product (*rpoD*-PCR.S: 5′-GCGAACTTACGTCTGGTTATT-3′; *rpoD*-PCR.AS: 5′-CGAACTGTTTACCCACTTCTT-3′; *rpoD*.Seq: 5′-CAGCTCGAGGGTGGTAT-3′) to identify recombinants that had acquired *rpoD* V445. Whole genome sequencing (detailed below) showed that the WT and variant strains were identical except for this one change.

The LF82 *rpoD* 445D mutant was constructed in a similar manner except we modified pRIPR by replacing the Cam^R gene with a gene cassette imparting kanamycin resistance (Kan^R) flanked by XbaI restriction sites and inserted into pRIPR digested with the same enzyme. Kan^R transformants were selected in competent EC100D cells (LGC Biosciences, Middlesex, UK), yielding pRIPR-Kan. The region of *rpoD* corresponding to nucleotide position 1334 and flanked by ∼500-bp of homology was obtained by PCR amplification of genomic DNA from LF82 with the *rpoD D445* genotype using primers *rpoD*-BspQI.S and *rpoD*-BspQI.AS and inserted into pRIPR-Kan at Nt.BspQI restriction sites. The resulting plasmid was transferred by conjugation into LF82, essentially as described above, with exoconjugants selected on LB plates supplemented with Kan (30 μg/ml). As LF82 is Amp^R, we also modified the helper plasmid pSceH by inserting a spectinomycin-resistant ($Spec^R$) gene cassette flanked by SmaI sites and inserting it into the plasmid digested with FspI (New England Biolabs). Recombinants were tested by Sanger sequencing of PCR products to identify the correct LF82 mutant.

Genome sequencing of MG1655 and MG1655^{D445V} strains

DNA was extracted from cells grown in LB media at 37° C in a shaking incubator at 250 rpm. The DNeasy Blood & Tissue kit (Qiagen, Beverly, MA, USA) was used for total DNA purification using the protocol for gram-negative bacteria. SMRTbell library preparation from microbial (genomic DNA) gDNA and whole genome sequencing using the Pacific Biosciences Sequel II platform (PacBio) were performed by the NCI CCR Sequencing facility. Genome sequencing data from this study are deposited with NCBI BioSample (Accession numbers: SAMN32661817-SAMN32661824) and summarized in Supplementary Table S1.

LF82 colonization into ASF mice

All animal protocols were conducted following approval of Iowa State University's Institutional Animal Care and Use and Institutional Biosafety Committees. Immunocompetent C3H:HeN mice harboring ASF were bred under germfree conditions at Iowa State University College of Veterinary Medicine (16). Irradiated mouse chow and sterile water were given *ad libitum*. The parent generation (G_0) was inoculated by oral gavage with 10^8 CFUs of a 'murine naïve' (*i.e.* freezer stock) LF82 at 6–8 weeks of age. A subset of the mice was then treated with 2% (*w*/*v*) dextran sodium sulfate (DSS) (five days on, six days off, five days on) to induce colitis (25,26). After cessation of the DSS treatment, mice were paired for breeding, such that mice treated with DSS were bred together, and untreated mice were bred together for multiple generations (*i.e.* up to five generations), such that LF82 was vertically transferred from dam to offspring. For each generation, LF82 was isolated from several mice on lactose MacConkey agar. Isolated colonies representative of the LF82 community were chosen and used to inoculate overnight cultures in 5 ml LB broth, which were then frozen in 80% glycerol at −80°C.

Genome sequencing of LF82 isolates recovered from ASF mice

LF82 isolates from the G1, G3 and G5 generations, with and without DSS, as well as a 'murine naïve' control, were grown in LB broth to approximately mid-log phase. gDNA was isolated using the MasterPure™ Complete DNA and RNA Purification Kit (Epicentre/Lucigen, Middleton, WI, USA) and the DNA purification protocol for bacterial cell samples. gDNA was quantified using the Nanodrop and Qubit 2.0 or 3.0 fluorometer with the broad range dsDNA kit.

Sequencing libraries for each isolate were created using the Nextera XT Sample Preparation Kit (Illumina, San Diego, CA, USA) using the standard workflow. Resulting libraries were then sequenced on the Illumina MiSeq at Iowa State University's DNA Facility.

Sequences for each isolate were aligned with the Burrows-Wheeler Aligner (27) using the reference genome for LF82 from NCBI (Accession No. NC_011993.1) (28). Reference sequence dictionaries were created using PicardTools. Duplicate removal, local realignment of indels, base quality score recalibration, variant calling, and hard filtering of variants were completed using GATK best practices (29–31). Because we were only interested in variants that occurred in the mouse gastrointestinal tract, any SNPs identified in the naïve control prior to hard filtering were used as known sites for the base quality score recalibration for the remaining isolates. Alignment statistics and quality were determined using Samstats (32). Hard filtered SNPs were filtered further using variant call file (VCF). SNPs were removed if the genotype (GT) was classified as heterozygous or homozygous to the reference or if the conditional genotype quality (GQ) was <99. Select SNPs that passed filtering were validated using standard end-point PCR methods on other isolates from the same generation. The resulting PCR products were sequenced at Iowa State University's DNA Facility. The sequences were mapped back to the reference genome and visualized using Geneious [Geneious version 10, www.geneious.com (33)]. Select SNPs were

compared to other *E. coli* sequences using NCBI's BLAST nucleotide alignment suite (34). Shotgun sequence data have been deposited under BioProject ID PRJNA912691; data are summarized in Supplementary Tables S2 and S3.

Analysis of σ^{70} D445 and σ^{70} V445 conservation

Protein fastq sequence files of σ^{70} were accessed and retrieved from NCBI using NCBI's EDirect tools. Protein sequences were blasted against WP_000437381.1 (the σ^{70} sequence of LF82) using the NCBI BLAST+tool. Basic command lines such as grep or awk were used to search for sequences that had a valine or an aspartic acid at the residue corresponding to the position equivalent to 445 in σ^{70} of LF82.

DNA

Construction of the pETσ^{70}*rpoD* D445V plasmid, for production of the σ^{70} D445V mutant protein, was done using the Q5® Site-Directed Mutagenesis Kit (NEB), following their protocol with recommended annealing temperature and pETσ^{70} CFI, which contains a His_6-tagged WT *rpoD* cloned into pET21(+) (35). Primers were designed using NEB's online primer design software, NEBaseChanger™ (https://nebasechanger.neb.com), to introduce the GAT to GTT mutation. After the kinase, ligase, and DpnI (KLD) Treatment (Step II, according to manufacturer's protocol), the KLD reaction mixture was transformed into BL21(DE3)/pLysE competent cells. Sanger sequencing (Macrogen, Rockville, MD, USA) verified the mutation.

Plasmids used for *in vitro* transcriptions are detailed in Supplementary Table S4. pJN1, which contains P_{S17}, and the plasmids containing LF82 promoter sequences were constructed by GenScript (Piscataway, NJ, USA) by inserting the indicated sequences within the multi-cloning site and upstream of the factor-independent, early transcription terminator of T7 present in the vector pTE103 (36). The expected size of the RNA from each promoter is indicated in Supplementary Table S4. All these plasmids also expressed RNAI from the promoter P_{RNAI}, which yields an RNA of 100–110 nucleotides (37).

Proteins

Core RNAP was purchased from NEB. WT σ^{70} and the σ^{70} D445V variant were purified as previously described [(38) and (39), respectively].

RNA-seq analyses

RNA was isolated from cells grown in LB media at 37° C or at 23° C in a shaking water bath at 250 rpm. Initial OD_{600} values were 0.1, and the cultures were grown until the OD_{600} reached a value between 0.5 and 0.6. The total RNA was isolated using Method II of Hinton (40). RNA was treated with DNase I (Turbo DNase, Life Technologies, Inc., Carlsbad, CA, USA) for 15 min at 37° C, and purified by phenol extraction/ethanol precipitation. RNA quality was assessed on a Bioanalyzer using the Agilent RNA 600 Nano Kit (Santa Clara, CA, USA). rRNA subtraction was performed to deplete ribosomal RNAs from each sample using the Ribo-Zero rRNA Removal Kit (Agilent Technologies, Gram-Negative Bacteria; Illumina, San Diego, CA, USA), and a TruSeq Kit mRNA Library Prep Kit (Illumina) was used to prepare the strand-specific cDNA library for Illumina sequencing. The library size was verified with a Bioanalyzer using an Agilent High Sensitivity DNA kit (Agilent Technologies). The concentration of each library was determined using the KAPA Library Quantification Kit (Roche, Basel, Switzerland) for Illumina platforms.

Sequencing was performed by the NIDDK Genomics Core facility using a MiSeq system with the single-end 50 bp Sequencing Kit (Illumina). Sequencing reads were quality-trimmed and remaining adapter sequences removed with the bbduk.sh tool from bbtools (v37.62). Next, paired reads were mapped to the GenBank *E. coli* LF82 GCF_000284495.1 or the *E. coli* MG1655 strain GCF_000005845.2 genome assemblies with STAR (v2.7.10a) (41), and quantification of fragments per gene was carried out with the featureCounts tool from Subread software (v2.0.1) (42) by counting reads falling within gene features. Differential gene expression analysis was performed with the R package DESeq2 (v1.42.0) (43). Estimated Log_2 Fold-Change values were adjusted with the lfcShrink function using the 'ashr' estimator (44). At least a 2.0-fold change in gene expression with an FDR-corrected P value of $\leq$0.05 was considered significant. Transcriptomic data are available in the NCBI database (GEO numbers GSE222248 and GSE253924) and in Supplementary Tables S5–S7.

RT-qPCR analyses

RT-qPCR was used to validate some of the RNA-seq data using the RNA isolated as described above, the *ssrA* gene (MG1655, accession ID: b2621) as an internal standard, and SsoAdvanced Universal SYBR Green Supermix (Bio-Rad Laboratories, Hercules, CA, USA) as a signal reporter. Reactions were performed in a 96-well microtiter PCR plate using 1 µl of cDNA at 1.5 ng/µl, and sense and antisense primers (0.5 µM each) were used to amplify each target gene in the Supermix. Primers for RT-qPCR were designed using Primer3web [https://primer3.ut.ee/] and purchased from Integrated DNA Technologies (Coralville, IA, USA; sequences are available upon request). Sample analysis was performed using CFX96 Touch Real-Time detection system (Bio-Rad). The cycling conditions were as follows: 98° C for 30 s (denaturation); 95° C for 30 s (amplification and quantification); 40 cycles of 95° C for 10 s, 52° C for 30 s, and 65° C for 30 s; a melting curve program of 65° C to 95° C with a heating rate of 0.5° C s^{-1} and continuous fluorescence measurement; and a cooling step to 65° C. The data were analyzed using the CFX manager software (Bio-Rad). The expression level of each sample was obtained by the standard curve for each gene and was normalized by the level of the internal control of the *ssrA* gene. Relative levels of gene expression were determined as a fold change (FC) using the threshold cycle $2^{-\Delta\Delta Ct}$.

For the statistical analyses, the mean FC and standard deviation (RT-qPCR mean ± std) were calculated using the 'mean()' and 'std()' functions, (respectively) from the Pandas library (https://pandas.pydata.org/). The standard error (SE) from the values for a particular gene was determined using the 'sem()' function from the SciPy Stats library. The 'ttest_1sampl()' function from the SciPy.Stats library (45) was used to conduct a one sample t-test to determine the *t* statistic and the *P* value.

Pathway analysis

Visualization of the transcriptomics data into representative categories was performed using a modified version of the EcoCyc Omics Dashboard tool (ecocyc.org) as described (46).

Each dataset was imported into an individual EcoCyc 'Smart Table' and the analysis was done using the Omics Dashboard Tool (47,48). A list of genes from each panel in the dashboard was downloaded and used to calculate the percentage of genes that were affected by the σ^{70} V445 versus D445 in the indicated strains using pandas v1.2.5 (49). Graphs were generated using Matplotlib graphics package in Python (50).

Growth curves

MG1655 and MG1655^{D445V} cells were streaked on 1.5% (*w*/v) LB (Lennox Broth) agar (Sigma-Aldrich, St. Louis, MO, USA) and grown overnight at 37° C. A single colony was collected, resuspended in LB (Quality Biological, Gaithersburg, MD, USA), and grown at 37° C with shaking at 250 rpm for 14–16 h overnight. Single colony stocks were stored in the presence of 50% glycerol at –80° C.

Overnight cultures of MG1655 and MG1655^{D445V} (obtained from a single colony or the stock) were diluted into 50 ml of the indicated media to a starting OD_{600} of 0.1 and then grown in a shaking (250 rpm) water bath at the indicated temperature. Growth curves (OD_{600}*vs.* time) were obtained in LB, EZ Rich Defined Medium lacking B12 and with or without 0.2 mM methionine (Teknova, Hollister, CA, USA), or simulated colonic environment medium (SCEM) (51).

SDS-PAGE of proteins made in EZ Rich Defined Medium with methionine was performed after growing cells to an OD_{600} of ∼0.55. Cells (1 ml) were harvested, the pellet was suspended in 1× Tris-tricine load solution (Thermo Fisher Scientific, Waltham, MA, USA), and the proteins were obtained after heating to 95° C and vigorous vortexing. The 10–20% Tris-tricine polyacrylamide gel (Thermo Fisher Scientific) was loaded with the equivalent of 0.08 OD_{600} units.

Streaking/cold sensitive phenotypes

To assess cold sensitive phenotypes, strains from frozen glycerol stocks were streaked on LB agar and incubated at 37° C for 24 h. The next day single colonies were selected, restreaked on LB agar and incubated at the following temperatures for the indicated time: 37° C (24 h), 30° C (24 h), 23° C (84 h), 16° C (132 h).

Biofilm formation

For biofilm formation, cells were incubated overnight at 37° C with shaking (250 rpm) in 5 ml of LB. The next day 2 μl of the overnight culture was inoculated in 198 μl of LB + 0.25% glucose in a 96-well plate (Corning, Corning, NY, USA). Plates were incubated without shaking at either 37° C (48 h) or at 23° C (96 h). Each well was washed once with 1× PBS (Quality Biological), allowed to dry for 5 min, and biofilm was stained with 200 μl of 0.1% crystal violet solution (Sigma-Aldrich) for 15 min followed by 2 washes of 1× PBS (200 μl each); after each wash wells were allowed to dry for 5 min. Biofilms were fixed by allowing them to completely dry for more than 24 h, and 0.2 ml 30% acetic acid (Macron Fine Chemicals/Avantor/VWR, Radnor, PA, USA) was used to solubilize the biofilm. To quantify the amount of biofilm, the absorbance was measured at 550 nm using a SpectaMax M3 plate reader (Molecular Devices, San Jose, CA, USA).

Immunoblots

An inoculum from an overnight culture (starting OD_{600} of 0.1) was grown in LB media at 37° C or at 23° C in a shaking water bath at 250 rpm. Once cells reached mid-exponential phase (OD_{600} between 0.5 and 0.6), 1 ml (for σ^{70} analysis) or 8 ml (for σ^{s} analysis) of cells were centrifuged for 10 min. Pellets were washed 2 times in 1× PBS. The pellet was resuspended in 1× Tricine sample buffer (2× Tricine buffer from Thermo Fisher Scientific) for a final concentration of 0.008 OD_{600} (samples for σ^{70}) or 0.032 OD_{600} (samples for σ^{s}) and heated at 95° C for 2 min with vigorous vortexing. A 1 μl aliquot of dilutions from the resuspensions was spotted onto a PVDF membrane (Thermo Fisher Scientific) presoaked with methanol, and the sample was allowed to absorb into the membrane. Membranes were blocked by incubating with 3% nonfat milk in 1× PBS overnight, then incubated for 1 h with σ^{s} or σ^{70} antibody (BioLegend, San Diego, CA, USA), as indicated, diluted in 0.3% nonfat milk in PBS-T [(1× PBS, 0.1% Tween-20 (Sigma-Aldrich)], and washed 4 times with 1× PBS-T. Blots were incubated for 1 h with the secondary antibody (HRP-Goat anti-mouse IgG, BioLegend), which was diluted in 0.3% nonfat milk in PBS-T. The blots were washed as before and then visualized using the ECL Plus Western Blotting Detection Reagent (GE Healthcare, Chicago, IL, USA) and detected using an OPTIMAX film processor (PROTEC, Oberstenfeld, Germany).

Western blot analyses

MG1655 and MG1655^{D445V} were grown to mid-log phase (OD_{600} ∼0.5–0.6, 'exp') and stationary phase (OD600 ∼3.00–4.00, 'stat') at either 37° C or 23° C in a shaking water bath. Western blots were performed as detailed below using standard methods, and protein amounts were normalized based on OD_{600} values. For RpoD westerns, 1 ml aliquots were centrifuged, and pellets were resuspended in 1× Laemmli sample buffer (Bio-Rad) for a final concentration of 0.008 OD_{600}. For the RpoS western blots, 8 ml aliquots were centrifuged, and the pellets were treated as follows such that the final concentration after the addition of Laemmli sample buffer was 0.032 OD_{600} per μl: Cells were incubated in a solution (∼80 μl for exp, ∼500 μl for stat) containing 10 mM Tris-Cl (pH 7.5), 1 mM EDTA, 10 mM DTT and lysozyme at 0.25 mg/ml (Sigma-Aldrich) for 10 min at room temperature (RT) and then lysed by 2 cycles of freeze [dry ice-ethanol bath)/thaw (RT water)]. A solution (3 μl for exp, 11 μl for stat) of 25 mM $MgSO_4$, 1.7 U/μl DNase (RNase-free, Thermo Fisher Scientific) was added, cells were incubated at 37° C for 15 min and placed on ice, and 2× Laemmli sample buffer containing 5% 2-mercaptoethanol (Sigma) was added. All samples were heated at 95°C for 2 min and vortexed vigorously until no pellet remained.

Proteins were separated by SDS-PAGE, and gels were dry blotted onto nitrocellulose membranes (iBlot 2 NC Ministacks, Invitrogen, Carlsbad, CA, USA) by the iBlot 2 transfer system (Invitrogen) at 20 V for 7 min. The membranes were blocked with 3% nonfat milk or ECL PrimeTM blocking agent (Cytiva, Marlborough, MA, USA) in PBS-T, washed three times with PBS-T (5 min each), and then incubated for 1 h with either anti-σ^{s} or anti-σ^{70} antibodies (Biolegend) in PBS-T containing 0.3% nonfat milk or ECL Prime^{TM} blocking agent. The membrane was washed three times (5 min

each) with PBS-T, then incubated with the secondary antibody (HRP-Goat anti-mouse IgG, BioLegend) in PBS-T containing 0.3% nonfat milk or ECL Prime™ blocking agent for 1 h at room temperature. After 3 washes with PBS-T (5 min each), the membrane was developed using the Amersham ECL Plus Western Blotting Detection System (GE Healthcare), and the signal was detected using an OPTIMAX film processor (Protec, Oberstenfeld, Germany).

Motility assay

A 1 μl culture aliquot (OD_{600} between 0.5 and 0.6) was dispensed in the middle of a 0.3% LB agar plate. After the liquid had absorbed into the surface, plates were incubated in the upright position at 37° C, 30° C, 23° C or 16° C, and the radius of the swarm was measured.

Minimum inhibitory concentration (MIC) assay

Isolated colonies from an overnight LB agar plate were resuspended in 1× PBS to achieve turbidity equivalent to a 0.5 McFarland standard. A sterile swab was dipped into the inoculum, excess liquid was removed, and the swab was streaked over an LB agar plate. The streaking was repeated 2 more times to ensure an evenly distributed lawn. After excess moisture was absorbed, a MTS™ MIC strip (Liofilchem, Waltham, MA, USA) was evenly applied onto the agar surface. Plates were inverted and incubated at 37° C for 12–14 h. MIC values were interpreted using The European Committee on Antimicrobial Susceptibility Testing (EUCAST) breakpoint tables and zone diameters (version 10.0, 2020) for the MIC strips. Assays were performed with at least two biological isolates.

In vitro transcription

Single round *in vitro* transcription reactions were performed by first reconstituting RNAP (0.05 pmol core: 0.2 pmol of the indicated σ) in a solution (1.9 μl) containing 1.28× transcription buffer [1× transcription buffer: 40 mM Tris-acetate (pH 7.9), 150 mM potassium glutamate, 4 mM magnesium acetate, 0.1 mM ethylenediaminetetraacetic acid (EDTA), 0.1 mM dithiothreitol (DTT), 100 μg/ml BSA] at the indicated temperature for 10 min. After addition of 3.1 μl containing 0.97× transcription buffer, 0.02 pmol pJN1 plasmid DNA and, when indicated, another promoter plasmid (0.02–0.05 pmol, as indicated in the figure), the solution was incubated at the indicated temperature for 10 min. Single round transcription reactions were initiated by the addition of a 1μl solution containing 1 mM ATP, CTP and GTP; 0.25 mM [α-^{32}P] UTP (~2 × 10^5 dpm/pmol); and 0.5 mg/ml heparin. After 10 min at the indicated temperature, reactions were collected on dry ice. RNA products were electrophoresed on 4% *w*/*v* polyacrylamide, 7 M urea denaturing gels as described (35). Gels were imaged by autoradiography followed by scanning with a GS-800 Calibrated Densitometer (Bio-Rad). Quantification was performed using Quantity One software (Bio-Rad).

Statistical analyses were performed using the number of independent transcriptions given in the figure legend and a t-test for 2 independent means (https://www.socscistatistics.com/tests/studentttest/default2.aspx).

Results

LF82 undergoes mutational changes in multiple genes during long-term colonization of ASF mice

LF82 fails to establish long-term colonization in mice with a complex microbiota, *i.e.* conventionally reared animals (17,52). Consequently, to identify genetic changes that might contribute to LF82 adaptation in a new host, we used ASF gnotobiotic mice that have a reduced microbial complexity, including the lack of *Proteobacteria* (53) since stable colonization of ASF mice using other bacterial species has been observed through multiple generations (16). As detailed in Materials and Methods, male and female young adult ASF mice were colonized once with a single dose of LF82 via oral gavage, and dextran sodium sulfate (DSS, 2.0%), which would induce inflammation and colonic epithelial cell damage, was either absent or present in the drinking water. Following each generation (1–5), aged breeder pairs were euthanized and necropsied. While no lesions or inflammation were observed in the GI tract epithelium of animals colonized with LF82, inflammation was observed in mice treated with DSS, consistent with prior studies (16). Cecal contents were plated after each mouse generation, and individual isolated LF82 colonies were whole-genome shotgun sequenced, along with the original WT LF82 strain, to identify SNPs that accrued over five generations.

As summarized in Supplementary Table S2, SNPs accumulated over time in both cohorts, and DSS treatment did not contribute to an increase in sequence changes. Alterations predicted to change amino acid sequences were found in genes associated with diverse cellular functions, including metabolite transport, metabolism and gene regulation. Unexpectedly, a sequence change in the highly conserved *rpoD* gene was repeatedly observed in mouse generation 3 (Supplementary Table S3).

LF82 has the novel σ^{70} sequence variant V445

rpoD encodes the σ^{70} subunit of RNAP, the primary 'housekeeping' sigma factor, which is needed for the expression of the majority of bacterial genes and for bacterial growth and viability [reviewed in (54)]. Although the σ^{70} sequence is highly conserved, in WT LF82 residue 445 is a valine (codon GTT) rather than the aspartate (codon GAT) that is found in >97% of primary σ's encoded by >20,000 *Enterobacteriales* sequences. A valine at this equivalent position is extremely rare (0.01%). It is not found in other AIEC (10–13,55–58), and BLAST identified only two additional *E. coli* genomes encoding σ^{70} V445. Both entries (HCO1463624.1 and HAM6750629.1) are nearly identical to LF82 σ^{70}. While these two strains are poorly characterized, genome sequence comparisons revealed that both appear to be distinct from LF82 and each other. Among species other than *E. coli*, a valine was also found at the equivalent position in the RpoD sequences of *Fusobacterium ulcerans* (SQJ00222.1) and *Fusobacterium varium* (EES64999.2). Interestingly, previous work has found *F. varium* in the mucosal microbiota of patients with UC (59).

Strikingly, after mouse colonization, the LF82 σ^{70} 'reverted' from the unusual V445 to the highly conserved D445. This occurred by a transversion at *rpoD* position 1334 converting the GTT codon to GAT (highlighted in light yellow in Supplementary Table S3). We speculate that the reversion to the conserved residue may have arisen because sta-

ble colonization of LF82 in the ASF gut allowed for a more 'commensal-like' existence for the human pathobiont.

σ^{70} V445 imparts new growth phenotypes to *E. coli*

To determine whether an aspartate or valine at σ^{70} residue 445 has any substantive effect on *E. coli*, we first introduced the *rpoD* D445V allele into the well-characterized *E. coli* MG1655 chromosome (generating MG1655^{D445V}). Whole genome sequencing indicated that this was the only difference between WT and the mutant allele (Supplementary Table S1). Growth comparisons of MG1655 versus MG1655^{D445V} in LB broth revealed similar exponential growth at 37° C, 30° C, 23° C, or 16° C (Supplementary Figure S1A). Similar growth kinetics were also observed using simulated colonic environment media (SCEM) (51) (Supplementary Figure S1B). Furthermore, the protein pattern observed by SDS-PAGE in exponentially growing cells at 37°C was similar in either strain (Supplementary Figure S1C). However, a modest growth defect at late exponential/stationary phase was seen with MG1655^{D445V}(Supplementary Figure S1A), and on plates MG1655^{D445V} was cold sensitive, with colonies displaying a distinctive mucoid phenotype as temperatures decreased (Supplementary Figure S1D).

Growth differences observed with σ^{70} V445 at 37°C are not attributable to changes in σ^{70} protein levels

In addition to σ^{70}, *E. coli* encodes 6 alternative σfactors (60). Thus, the growth change for MG1655^{D445V} at late exponential/stationary phase might arise from a difference in the steady-state levels of σ^{70} in the strains, which could affect sigma competition. Consequently, we performed western blots with σ^{70} antibody using cell lysates prepared from exponential and stationary phase cultures grown at 37°C. We found similar amounts of σ^{70} in MG1655 and MG1655^{D445V} (Supplementary Figure S2A; Supplementary Figure S3), indicating that the altered growth phenotype of MG1655^{D445V} does not result from altered cellular levels of σ^{70} at this temperature, and suggesting that it arises instead from altered function.

σ^{70} V445 affects expression of specific genes

To obtain insight into the effect σ^{70} D445 versus σ^{70} V445 on gene expression, we introduced the *rpoD* V445D allele into LF82, generating LF82^{V445D} and performed RNA-seq analyses using triplicate cultures of MG1655, MG1655^{D445V}, LF82, and LF82^{V445D} grown at 37°C (Supplementary Tables S5 and S6). Using a FC ≥2 and an adjusted *P*-value ≤0.05, we identified numerous genes whose RNA levels significantly increased or decreased in the presence of σ^{70} V445 (Table 1; Supplementary Tables S5 and S6). RT-qPCR analyses confirmed the results for selected genes in the MG1655^{D445V}*vs.* MG1655 datasets (Supplementary Table S5).

As an initial investigation of the transcriptomic data, we assigned cellular functions to the differentially expressed genes and looked for pathways where >50% of associated genes showed either increased or decreased FCs in either the MG1655 or LF82 backgrounds using the EcoCyc Omics Dashboard tool (ecocyc.org) (46) (Supplementary Figure S4). None of these pathways were affected except for one subpathway (toxin formation), which was down in WT LF82 versus LF82 V445D. This sub-pathway has only one gene, *cvpA*, which is required for colicin V production.

Table 1. Summary of RNA-Seq analyses

Comparison[a]	Temperature	Genes up	Shared	Genes down	Shared
MG1655^{D445V}*vs.* MG1655	37° C	73		33	
	23° C	192	51	306	18
LF82 *vs.* LF82^{V445D}	37° C	69	24	76	4

[a]RNA-seq analyses (Supplementary Tables S5–S7) were performed using RNA isolated from the indicated strains grown at the indicated temperature. Genes up and genes down had increased or decreased FCs ≥2, respectively, with an adjusted p-value of ≤0.05. Shared genes represent the number of genes in common with those in the MG1655^{D445V}*vs.* MG1655 at 37° C dataset.

MG1655 and LF82 are genetically disparate *E. coli* with dozens of SNPs between them (56). Thus, to identify genes whose expression might be directly related to σ^{70} V445, we looked for genes whose expression changed with σ^{70} V445 in both the MG1655 and LF82 backgrounds (Tables 2 and 3; Figure 1). We found 24 shared genes with FC increases, which represented a significant fraction (∼1/3) of the up genes in either dataset individually. Among these genes were those whose products were involved in specific cellular processes, such as methionine biosynthesis (*metE*), antibiotic resistance (*ampC*, encoding β-lactamase and the multi-drug efflux pumps *acrZ* and *bcr*), and degradation of L-fucose (*fucI*), as well as several genes with membrane-associated functions, such as transporters and pumps (*shiA*, *mlaA*, *cysZ* and *sppA*). We also found genes within the operon *lptM-dapF-yigA-xerC-yigB*, which encodes products of diverse functions. Finally, *yobH* was identified in both datasets. Although the function of YobH has not been characterized in *E. coli*, in *Salmonella* it has been shown to be involved in the invasion of macrophages (61).

Eight of the top 10 genes with increased expression in the LF82 dataset were exclusive to the σ^{70} V445 LF82 background (Table 2). One of these, the gene with the highest expression, was *RS00415*, annotated to encode a putative small protein, FruL. However, in the MG1655 genome the annotation has been removed and *fruL* has been classified as a 'phantom' gene that does not exist (62). Consequently, to investigate whether expression of this region also changes in MG1655, we manually annotated the gene in the MG1655^{D445V}*vs.* MG1655 analysis (Table S5). We did not find an increase, indicating that the *RS00415* FC is specific to LF82.

Unlike the significant overlap of genes found with increased expression, only 4 genes with decreased expression were shared between the datasets, representing 12% of the MG1655^{D445V} genes and 5% of the LF82 genes with lowered FCs, respectively (Tables 1 and 3).

σ^{70} V445 results in a different antibiotic resistance pattern

Ampicillin resistance is afforded by *ampC*, encoding β-lactamase. Although *ampC* is present on the chromosome of LF82 and other strains of *E. coli*, including MG1655, LF82 is known to be ampicillin resistant (Amp^R), while MG1655 is amp sensitive (Amp^S) (Table 4). Previous work has ascribed the Amp^R of LF82 to a mutation within the promoter for *ampC*, which increases promoter activity, together with a mutation within the downstream attenuator located in the 5′ untranslated *ampC* region, which decreases termination [(63); (DeWolf, *et al.*, manuscript in preparation)]. However, since

Table 2. Genes with increased FCs for σ^{70} V445

Gene	Function	MG1655^{D445V} *vs.* MG1655			LF82 *vs.* LF82^{V445D}	
		Rank[a]	FC[b]	RT-qPCR[c]	Rank[a]	FC[b]
Genes observed in both MG1655^{D445V} *vs.* MG1655 and LF82 *vs.* LF82^{V445D} datasets:						
metE	methionine biosynthesis	1	8.94	8.4 ± 1.2	46	2.28
yobH[d]	uncharacterized protein	3	4.32		12	3.46
lptM[e]	oxidative maturation associated lipoprotein	5	3.84		6	4.08
yigA	putative protein	8	3.27		7	3.97
xerC	recombinase	7	3.29		23	2.95
yigB	riboflavin biosynthesis	10	3.05		27	2.77
ampC	β-lactamase	6	3.32	3.4 ± 0.6	20	3.07
yciY	uncharacterized protein	9	3.27		13	3.46
shiA	shikimate:H+symporter	13	2.99		30	2.66
acrZ	efflux pump accessory	14	2.89		17	3.27
slt	murein transglycosylase	18	2.79		41	2.35
fucI	L-fucose isomerase	19	2.77	2.0 ± 0.2	43	2.31
ybdG	mechanosensitive channel	22	2.66		15	3.43
mlaA	outer membrane lipoprotein	30	2.51		31	2.60
cysZ	sulfate:H+symporter	31	2.48		52	2.16
dpaA	peptidoglycan meso-diamino acid amidase A	33	2.46		28	2.73
xseA	exoDNase VII subunit	35	2.43		21	3.05
mepK	peptidoglycan endopeptidase	47	2.33		35	2.53
aroP	aromatic aa:H^+ symporter	54	2.22		14	3.43
mnmC	tRNA 5-aminomethyl-2-thiouridylate synthase	60	2.14		33	2.58
rsuA[e]	pseudouridine synthase	61	2.10		25	2.87
bcr	multidrug efflux pump	66	2.06		65	2.03
csrD	regulator of CsrB/CsrC	72	2.00		47	2.28
sppA	signal peptide protease	73	2.00		37	2.43
Top increased genes observed exclusively in MG1655^{D445V} *vs.* MG1655						
ymgG	PF13436 family protein	2	4.99			
RprA	sRNA, σ^s regulation	4	4.20			
yggE	ROS response	11	3.03			
hdeA	acid resistance	12	2.99	4.1 ± 0.7		
hdeB	acid resistance	15	2.85	4.5 ± 0.9		
nanA	degradation of sialic acid	16	2.85			
slp	starvation lipoprotein	17	2.81			
AceK-int	sRNA	20	2.75			
Top increased genes observed exclusively in LF82 *vs.* LF82^{V445D}						
RS00415	phantom gene *fruL*				1	7.94
ygbJ	putative L-threonate dehydrogenase				2	4.20
nhoA	arylamine *N*-acetyltransferase				3	4.17
ygbL	putative decarboxylase				4	4.17
lacZ	β-galactosidase				5	4.11
xdhC	putative xanthine dehydrogenase				8	3.86
deoA	thymidine phosphorylase				9	3.84
deoC	deoxyribose-phosphate aldolase				10	3.76
ygbK	putative 3-oxo-tetronate kinase				11	3.61
tsx	nucleoside-specific channel forming protein				16	3.29
betI	DNA-binding transcriptional repressor				18	3.25
dapF	diaminopimelate epimerase				19	3.18

[a]Ranked according to fold change in list of up-regulated genes in Supplementary Tables S5 and S6.
[b]Fold change in RNA-seq analyses.
[c]Where a value is indicated; fold change was determined using the threshold cycle $2^{-\Delta\Delta Ct}$.
[d]Genes in bold have a promoter whose activity increased when using σ^{70} V445 versus σ^{70} D445 in *in vitro* transcription (Figure 4).
[e]*lptM* to *yigB* compose the operon *lptM-dapF-yigA-xerC-yigB; rsuA* and *bcr* compose the operon *rsuA-bcr.*

the FC for *ampC* was elevated in both the MG1655^{D445V} versus MG1655 and the LF82 *versus* LF82^{V445D} datasets (Table 1; Supplementary Tables S5 and S6), the potential thatσ^{70} V445 might also contribute to the Amp^R phenotype of LF82 was tested. For this, we performed minimum inhibitory concentration (MIC) analyses for variousβ-lactam antibiotics (Table 4). MG1655^{D445V} exhibited significantly increased resistance relative to MG1655 for all fiveβ-lactam antibiotics tested. For example, the MIC for ampicillin was 4μg/ml for MG1655 and 12μg/ml for MG1655^{D445V}. Given that *E. coli* with an MIC greater than 8μg/ml is considered Amp^R (64), we conclude that theσ^{70} D445V mutation imparts Amp^R to MG1655.

Both LF82 and LF82^{V445D} were highly resistant to ampicillin, showing no sensitivity even at the highest concentration (Table 4). Given the presence of the mutated promoter/attenuator region in the LF82 genome, this was not particularly surprising. Consequently, we tested LF82 and LF82^{V445D} with anotherβ-lactam antibiotic, cefoxitin, which was more potent with MG1655. In this case, we observed a

Table 3. Genes with significant decreased FCs in both MG1655 and LF82 datasets

		MG1655^{D445V} versus MG1655		LF82 versus LF82^{V445D}	
Gene	Function	Rank[a]	FC[b]	Rank[a]	FC[b]
gltP	glutamate/aspartate : H+symporter	12	−2.28	53	−2.23
folD	Methylenetetrahydrofolate de/cyclo hydrogenase	22	−2.10	70	−2.08
ruvA	Holliday junction binding subunit	23	−2.08	65	−2.11
yacC	putative lipoprotein	32	−2.01	66	−2.10

[a]Ranked according to fold change in list of down-regulated genes in Supplementary Tables S5 and S6.
[b]Fold change in RNA-seq analyses.

decrease in LF82 resistance withσ^{70} D445, consistent with the results seen with MG1655 *vs.* MG1655^{D445V} (Table 4).

Susceptibility to antibiotics is also affected by various multi-drug efflux pumps, such as those encoded by *acrZ*, *bcr*, and *mdtk* (65,66), whose expression increased withσ^{70} V445 in the MG1655 background (Table 2, Supplementary Table S5). In addition, increased expression of *acrZ* and *bcr* was also seen withσ^{70} V445 in the LF82 background (Table 2; Supplementary Table S6). Consequently, we tested MG1655 and MG1655^{D445V} in MIC assays for resistance to additional antibiotics, including known substrates for these pumps (65,66) [chloramphenicol and norfloxacin (AcrZ, MdtK), kanamycin (Bcr), tetracycline (AcrZ, Bcr) and ciprofloxacin (MdtK)] along with gentamicin and streptomycin. MG1655^{D445V} showed increased resistance to streptomycin, kanamycin, gentamicin and ciprofloxacin, reduced resistance to chloramphenicol and tetracycline, and no difference for norfloxacin (Table 4), indicating thatσ^{70} V445 imparts an overall change in the antibiotic resistance pattern.

σ^{70} V445 results in better growth in the absence of methionine

MetE catalyzes the final step of methionine biosynthesis in the absence of cobalamin (vitamin B12) [reviewed in (67)]. In both the MG1655 and the LF82 analyses, the level of *metE* RNA increased by the presence ofσ^{70} V445 (Table 2; Supplementary Tables S5, S6). In fact, among the MG1655 genes, *metE* showed the highest FC increase in the entire 37° C dataset. Consequently, we investigated the growth kinetics of MG1655 and MG1655^{D445V} in a defined rich medium (EZ Rich) lacking B12 with and without methionine. While MG1655 grew better in the presence of methionine, MG1655^{D445V} grew better in the absence of methionine (Figure 2, grey *vs.* blue lines, respectively). Although the difference was modest, this result suggests that σ^{70} V445 increases the ability to synthesize methionine, imparting a growth advantage when exogenous methionine is limited.

σ^{70} V445 results in increased biofilm and a dysregulation of σ^{s}

LF82 is known to produce more biofilm than commensal *E. coli*, and this ability has been shown to be a phenotype that distinguishes AIEC from other *E. coli* strains (68). To investigate whether MG1655^{D445V} acquired this characteristic, we performed crystal violet staining assays. We found that biofilm formation was significantly higher in the MG1655^{D445V} strain compared to MG1655 at both 37° C and 23° C (Figure 3).

Alterations in the level of σ^{s}, the alternative sigma factor used during stationary phase and during certain times of stress (60), are known to affect the level of biofilm (69). Regulation of σ^{s} is complex, involving multiple transcriptional and post-transcriptional controls (60,70). No significant FC differences in *rpoS* RNA were observed for σ^{70} V445 in either the MG1655 or LF82 backgrounds. However, growth of MG1655^{D445V} in broth was modestly impaired relative to MG1655 at higher cell density (Supplementary Figure S1A), a result that could arise from a decrease in σ^{s} amount or activity. Western and dot blots confirmed that the steady-state levels of σ^{s} in exponential and stationary MG1655^{D445V} cultures at 37° C were ∼2-fold lower relative to MG1655 cultures (Supplementary Figure S2B, Supplementary Figure S3B). As expected, extensive σ^{s} proteolysis was observed in exponential cultures, where various factors are known to promote σ^{s} degradation (70). This result suggests that σ^{70} V445 indirectly affects the amount of σ^{s} through one or more of the post-transcriptional pathways.

Dysregulation of σ^{s} is not responsible for the gene expression differences observed with σ^{70} V445

Hundreds of genes are differentially regulated by σ^{s}, either directly or indirectly (71). Thus, the FCs we observed in our RNA-seq analyses might be an indirect consequence of σ^{s} dysregulation. However, there was practically no correlation between the affected genes in our RNA-seq analyses (Supplementary Tables S5, S6) and a previous RNA-seq analysis of the *rpoS* regulon (71). Of the 24 genes, whose FCs increased with σ^{70} V445 in both the MG1655 and LF82 backgrounds (Table 2), only one, *fucI*, which is negatively regulated by RpoS, could potentially have increased because of a decrease in the level of σ^{s}. Of the 4 genes with decreased FCs with σ^{70} V445 in both the datasets, none were positively regulated by σ^{s}.

RNAP containing σ^{70} V445 is more active at specific promoters *in vitro*

To investigate whether the σ^{70} D445V variant directly changes the ability of RNAP to initiate transcription at particular promoters, we performed *in vitro* transcriptions using RNAP containing σ^{70} D445 or σ^{70} V445. We first used a σ^{70}-dependent promoter (P_{S17}) that has high sequence similarity to the consensus -10 and -35 sequence motifs and an ideal spacer length of 17 bp at 37° C (Figure 4). Single round transcription was used to observe changes specifically during transcription initiation.

We found no significant difference between RNAP containing σ^{70} D445 or σ^{70} V445. Using P_{S17}, we also separated the process of RNAP reconstitution from transcription by reconstituting σ^{70} with core at either 30° C or 37° C and then incubating RNAP with the supercoiled DNA at either 30° C or 37° C before the start of transcription (Supplementary Figure S5). We chose different temperatures since MG1655 and MG1655^{D445V} exhibit different growth characteristics on plates at 30° C versus 37° C (Supplementary Figure S1D). Again, we observed similar activities for σ^{70} D445 and σ^{70} V445.

To investigate if the σ^{70} mutation affected transcription from specific genes, we used the Ecocyc Genome Database (47) to inspect the genomic organization of the genes whose

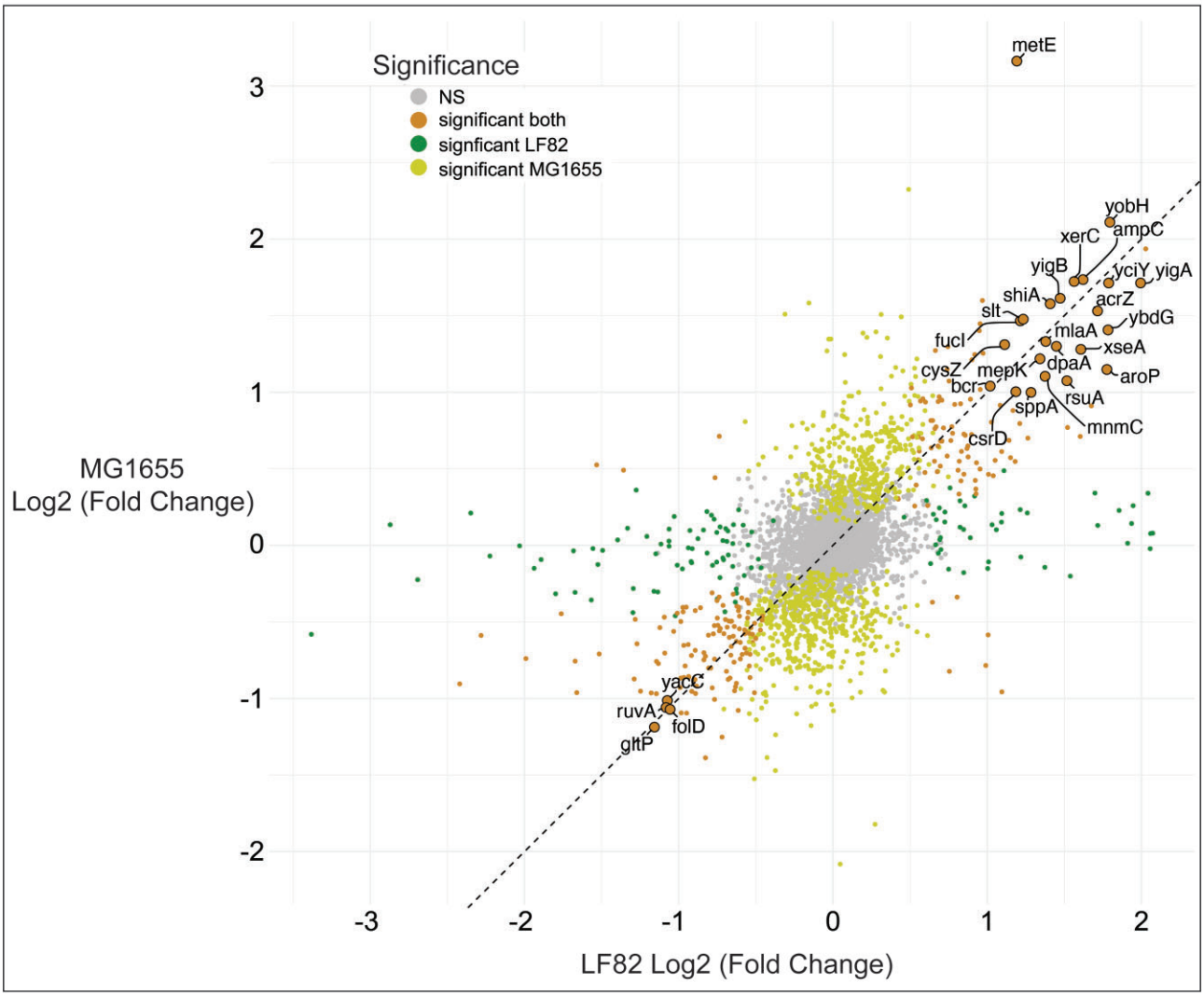


Figure 1. Comparative analysis of the impact of σ^{70} V445 on the transcriptomes of LF82 and MG1655. The two axes depict the Log_2 fold change in gene expression between LF82 and LF82^{V445D} (X-axis) and between MG1655^{D445V} and MG1655 (Y-axis). Orange indicates genes significantly perturbed in both strains; those genes with an adjusted p-value $\leq$0.05 and |Log_2 fold change| $\geq$1 are larger represented by the larger, labeled circles. Genes that only change in either the MG1655 or LF82 backgrounds are highlighted in yellow or green respectively, while grey represents genes that are not affected by σ^{70} V445.

transcripts increased with σ^{70} V445 in both the MG1655 and the LF82 backgrounds (Table 2). We found that most of these genes were either 'stand alone', *i.e.* a single gene, or present within operons (*lptM-dapF-yigA-xerC-yigB* or *rsuA-bcr*) immediately downstream of a predicted or identified σ^{70}-dependent promoter (Supplementary Table S4). Consequently, we investigated some of these promoters in the *in vitro* transcription assay. We used supercoiled templates containing regions of LF82 DNA upstream of *yobH*, *ampC*, *yciY*, *xseA*, *csrD*, *sppA* and *lptM*, genes whose RNA increased in both the MG1655 and LF82 backgrounds (shown in bold in Table 2) as well as *RS00415* and *betI*, genes whose RNA increased only in the LF82 background (Figure 4A). Our transcription reactions also contained P_{S17}, as a control, as well as the promoter for RNAI, P_{RNAI} (37), which was present on the plasmids. We determined the expected transcript sizes for the tested promoters based on previous promoter identifications or by predicting possible promoter(s) within the inserted DNA sequence (Supplementary Table S4).

P_{S17}, P_{RNAI}, $P_{RS00415}$ and P_{betI} were similarly active with either σ^{70} D445 or σ^{70} V445 (Figure 4B, C), indicating that σ^{70} V445 did not affect transcription at these promoters. In contrast, all of the tested promoters for genes whose expression increased in both the MG1655 and LF82 backgrounds (P_{ampC}, P_{lptM}, P_{yobH}, P_{xseA}, P_{yciY}, P_{csrD} and P_{sppA}) were ~2- to 3-fold more active with σ^{70} V445 than with σ^{70} D445. It should be noted that the activities of the tested promoters varied relative to P_{S17}, from stronger (P_{xseA}) to weaker (P_{yobH} and P_{RNAI}). Thus, promoter activity *per se* did not correlate with a preference for RNAP with σ^{70} V445.

Since these reactions were performed as single-round transcriptions, the increase in transcription must reflect an increase in transcription initiation, rather than elongation or termination. Thus, we conclude that for *ampC*, *yobH*, *xseA*, *yciY*, *csrD*, *sppA* and the *lptM* operon, σ^{70} V445 is, at least in part, directly responsible for the FC increases seen at these genes *in vivo*, by affecting a step(s) during the initiation of transcription.

Table 4. Minimum Inhibitory Concentration (MIC) for MG1655 and MG1655^{D445V} with various antibiotics

Antibiotic	Type	MIC (μg/ml)	
		MG1655	MG1655^{D445V}
	β-lactam		
Amoxicillin		4	**14**[a]
Ampicillin		4	**12**
Cefoxitin		2.3	**3.3**
Cephalothin		10	**28**
Penicillin G		52	**>256**
	Protein synthesis inhibitor		
Chloramphenicol	Aminoglycoside	8	**5**
Gentamicin	Aminoglycoside	0.25	**0.5**
Kanamycin	Aminoglycoside	0.625	**2**
Streptomycin	Aminoglycoside	0.5	**1.5**
Tetracycline	Tetracycline	0.625	0.44
	Replication inhibitor		
Ciprofloxacin	Quinolone	0.016	**0.0195**
	Topoisomerase inhibitor		
Norfloxacin	Fluoroquinolone	0.064	**0.064**
Antibiotic	**Type**	**LF82^{V445D}**	**LF82**
	β-lactam		
Ampicillin		>256	>256
Cefoxitin		16	**96**

[a]Numbers in bold are for those cases in which the presence of σ^{70} V445 resulted in more resistance.

A 16 bp spacer and a G:C at promoter position −14 correlate with increased transcription by RNAP containing σ^{70} V445

In the absence of an activator, recognition of the −10 and −35 elements by σ^{70}-RNAP requires a spacer length of 16–18 bp, with 17 bp being ideal [reviewed in (18)]. Since by convention the −10 element is located from positions −12 to −7, this places the spacer from position −13 to position −29 (±1 bp). Sequence analysis of the promoters with increased *in vitro* transcription when using RNAP with σ^{70} V445 *vs.* σ^{70} D445 indicated that 5 out of 7 of the promoters with increased activity shared a 16 bp spacer with a G:C at position −14 (Figure 4A), suggesting that this property might be related to the ability of σ^{70} V445 to increase initiation at these promoters. However, this property was not absolute as both P_{crsD} (16 bp spacer) and P_{sppA} (17 bp spacer) had a −14 A:T, yet demonstrated 2.6- and 2.3-fold increases, respectively.

To investigate this correlation further, we inspected the promoters used in our *in vitro* transcription together with the σ^{70}-dependent promoters listed within the Ecocyc Genome Database (47) for all of the genes that exhibited a significant increase with σ^{70} V445 in both the MG1655 and LF82 backgrounds (Table 2). As listed in Supplementary Table S4, we found 17 promoters. We excluded the *shiA* promoter PM0-46179 since *shiA* is also expressed by PM0-46175, which, unlike PM0-46179, has been validated by experimental evidence. We also excluded the P_{aroP} promoters since their sequences are not conserved in LF82. The total list (marked in green in Supplementary Table S4) had 14 promoters, including the 7 that were tested *in vitro* (Figure 4). Among the 14 promoters, nine had a predicted 16 bp spacer (64%), ten had a predicted -14G:C (71%), and eight had both features (57%). Previous analyses of hundreds of *E. coli* promoter sequences (72,73) found that 20–25% have a G:C at position −14, ∼20% have a 16 bp spacer, and ∼5% have both. Thus, our finding that >50% of the promoters associated with increased expression have the 16 bp spacer/−14G:C combination far exceeds random chance.

To investigate how the bp at −14 and the 16 bp spacer affected transcription of one of these promoters *in vitro*, we tested P_{ampC} with various mutations (Figure 5). We found that replacing the P_{ampC} spacer with that of P_{S17} ($P_{ampC,S17}$), a 17 bp spacer with a T:A at -14, eliminated the preference by σ^{70} V445-RNAP, indicating that at this promoter, the spacer imparted the preference. We also tested the effect of each possible bp at −14 (P_{-14T}, P_{-14A}, P_{-14C}). Each of these was still more active when using σ^{70} V445 *vs.* σ^{70} D445, indicating that for P_{ampC} it is the presence of a 16 bp spacer, rather than the bp at position −14, that is responsible for σ^{70} V445 preference.

We then tested a set of mutations that increased the spacer length to 17 bp, by inserting a bp at −14, −16 or −18 (P_{+14T}, P_{+16G}, P_{+18A}, respectively), or by both changing the −14 G:C to T:A and increasing the spacer to 17 bp ($P_{+18A/-14T}$, $P_{+16G/-14T}$). All of these mutations resulted in a higher overall activity for the promoter. This is expected since 17 bp is the ideal spacer length. Importantly, each promoter was now similarly active using either σ^{70} V445 or σ^{70} D445. Taken together these results suggest that in the case of P_{ampC}, it is the 16 bp spacer, regardless of the bp at −14, that increases the activity of the promoter with RNAP containing σ^{70} V445. However, it should be noted that we estimate that we would need at least a 50% effect in this assay in order to reproducibly observe a change. Consequently, we are not able to quantify subtle effects.

σ^{70} V445 results in a wider range of effects at 23° C than at 37° C

Although MG1655 and MG1655^{D445V} grew similarly in exponential liquid cultures and on plates at 37° C, more distinct differences in growth were observed on plates at lower temperatures (Supplementary Figure S1D). To further characterize the cold-sensitive growth phenotype of MG1655^{D445V} and to gain insights into how the bacteria might respond to lower temperatures as cells are shed from the mammalian gut, we also performed RNA-seq analyses comparing MG1655 *vs.* MG1655^{D445V} at 23° C. We found that ∼27% of the genes with increased FCs at 23° C were also observed at 37° C (Table 1; Supplementary Tables S5, S7). Of the 24 genes with increased expression in both the MG1655 37° C and LF82 datasets (Table 2), only 2 (*shiA* and *mepK*) were not seen in the MG1655 background at 23° C (Supplementary Table S7). The up genes included those related to changes in antibiotic resistance patterns (*ampC*, *acrZ* and *bcr*) and methionine biosynthesis (*metE*). In addition, though, a large number of genes with increased expression at 23° C were exclusive to that temperature. These included the most increased genes, the *fruB-fruK-fruA* operon, which is important for fructose utilization, and all 8 genes within the pathway for ATP synthesis (*atpA-H*) (Supplementary Table S7; Supplementary Figure S4).

About half of the genes with decreased FCs at 37° C were also seen at 23° C. However, there were 288 genes exclusive to the lower temperature. These included >50 genes involved in flagella formation, chemotaxis, and motility (Supplementary Table S7). Consequently, >95% of flagellar genes were significantly down with MG1655^{D445V} versus MG1655 at 23° C (Supplementary Figure S4). In contrast, at 37° C only *flhA*, which is necessary for flagellin export, decreased significantly.

To investigate whether the decreased FCs observed in multiple motility genes at 23° C and in *flhA* at 37° C were consequential, we performed plate-swarming mobility assays. Consistent with the gene expression results, MG1655^{D445V} was

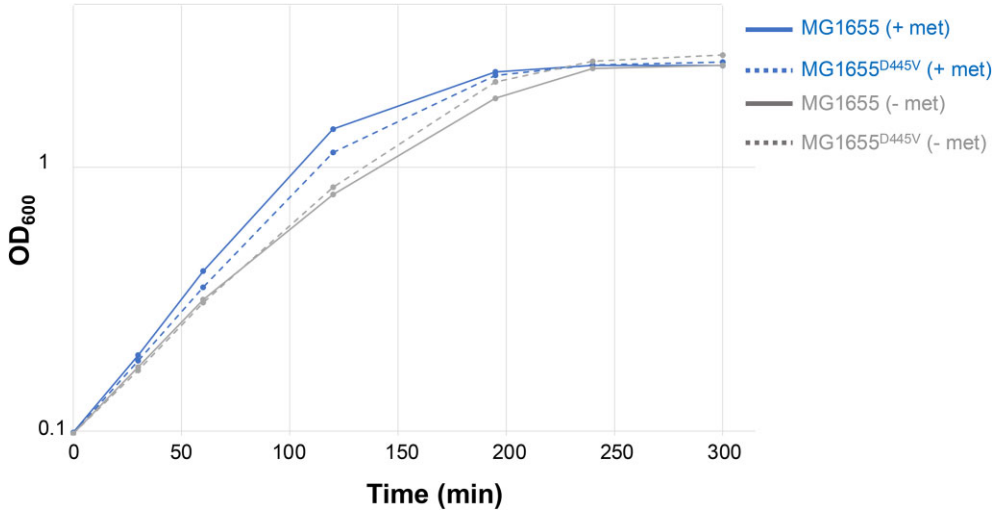


Figure 2. MG1655^{D445V} grows better than MG1655 in the absence of methionine. Growth curves (representative of two independent experiments) show OD_{600} (in log plot) *vs.* time for MG1655 (solid line) or MG1655^{D445V} (dotted line) grown in EZ Rich Defined Media lacking B12 and without (grey lines) or with (blue lines) 0.2 mM methionine (met).

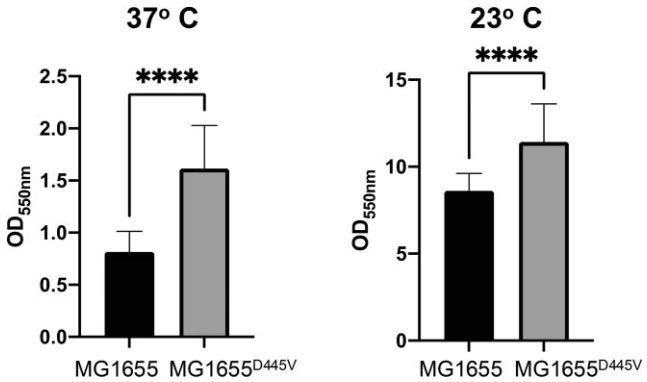


Figure 3. Level of biofilm observed at OD_{550} after crystal violet staining of MG1655 or MG1655^{D445V} cultures grown in LB at 37°C (left) or 23°C (right). Means, standard deviations, and two sample *t*-tests performed in R were generated from values obtained in a 96-well plate with 24 replicates for each strain (**** *P*-value $< 10^{-8}$).

less motile than MG1655 as temperatures decreased from 37° C to 23° C to 16° C (Supplementary Figure S6). Given that MG1655^{D445V} and MG1655 form similar sized colonies on plates at 37° C (Supplementary Figure S1D), we conclude that the σ^{70} V445 variant impairs motility at 37° C. At the lower temperatures, motility is impaired either directly or it occurs indirectly by the decreased growth rate of MG1655^{D445V} relative to MG1655.

We also performed western blot analyses to investigate whether σ^{70} or σ^{s} levels were affected by σ^{70} V445 at 23° C. While σ^{70} increased 1.5- to 2-fold, σ^{s} decreased ∼2-fold when comparing MG1655^{D445V} to MG1655 at this temperature (Supplementary Figure S2, S3). These changes in the steady state levels of σ^{70} and σ^{s} may be at least partly responsible for the increased number of affected genes at 23° C *vs.* 37° C. However, as was seen with the affected genes at 37° C, there was very little overlap between the 23° C σ^{70} D445V regulon and the *rpoS* regulon (Supplementary Table S7).

Taken together these results indicated that at 23° C, σ^{70} V445 results in a wider range of gene expression changes than seen at 37° C. Given the effect on σ^{70} levels, it is likely that both the σ^{70} function and the amount of σ^{70} V445 play a role in the various expression changes.

Discussion

The single D445V substitution within σ^{70} affects the expression of specific genes and leads to altered phenotypes

How *E. coli* evolved to occupy host-specific niches is highly relevant to inflammatory diseases of the gut, where *E. coli* pathobionts are associated with the emergence and severity of disease (3). Studies in *E. coli* reveal the importance of the host gut environment in promoting genetic diversity and the emergence of new phenotypes within the species (14,74), where even relatively few differences in genetic composition can result in phenotypic differences that contribute to selective advantages (75–77). However, despite extensive efforts, including whole genome and metagenomic sequencing, until this work no genetic features have been identified within AIEC strains, such as LF82, that can account for their phenotypic differences from 'commensal' or pathogenic *E. coli* (11,12,55,78).

Efforts to identify genetic factors that contribute to host adaptation have been hampered by the lack of robust mouse models that can establish long-term colonization of AIEC strains (17,52). Consequently, we colonized defined microbiota ASF mice with LF82 since with this model we could identify genotypic changes that occurred over five generations.

Unexpectedly, our SNP analysis revealed that during LF82 transmission through ASF mice, a rare substitution within the highly conserved σ^{70} subunit of RNAP 'reverted back' to the conserved D445. We hypothesize that the unique environmental conditions in the gut of CD patients (20) originally selected for the emergence of the *rpoD* variant, while adaptation within the ASF gut, which lacks other proteobacteria, allowed for a more 'commensal-like' existence, resulting in the reversion.

Our investigation of RpoD D445 versus V445 indicated that this one single change alters the activity of RNAP containing σ^{70} V445 at specific promoters, which in turn affects

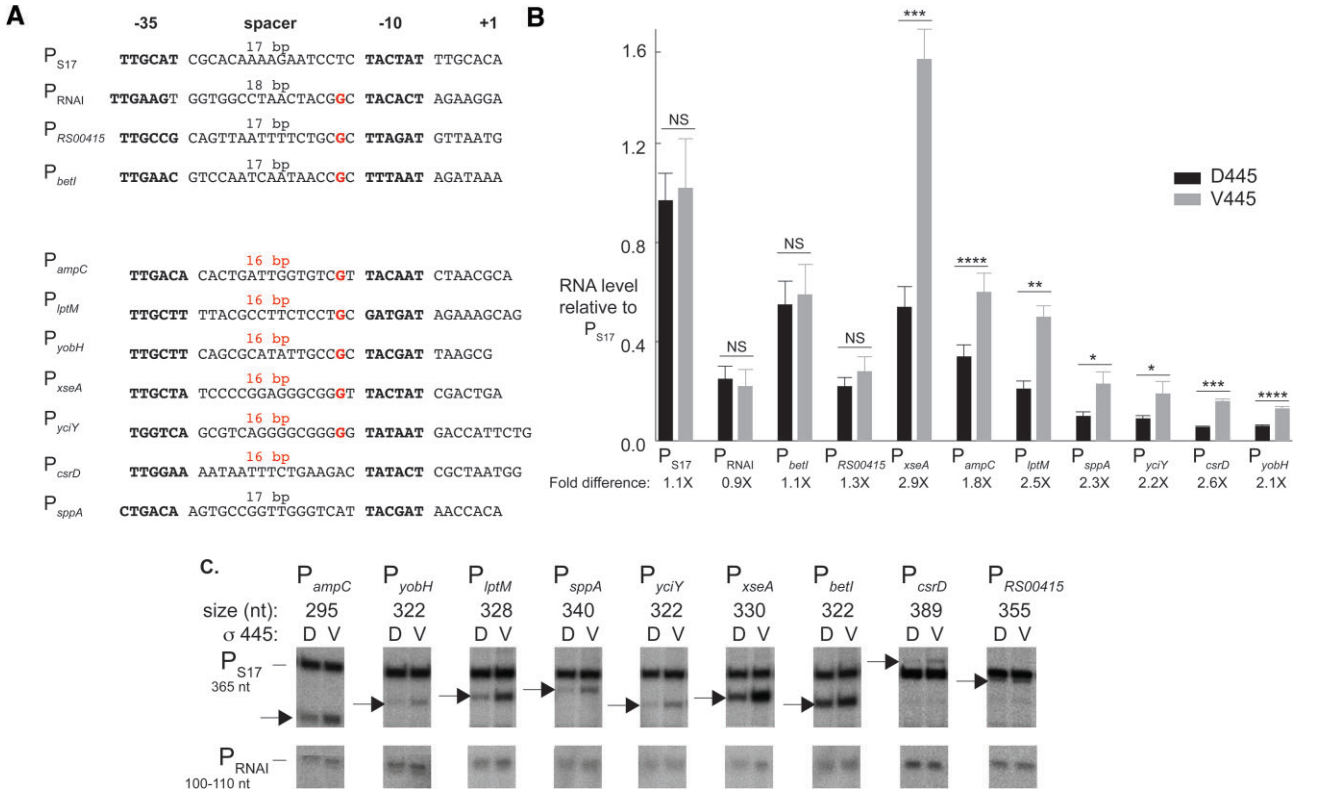


Figure 4. Specific promoters are more active with σ^{70} V445-RNAP than with σ^{70} D445-RNAP in *in vitro* transcriptions. (**A**) Sequences of tested promoters from the -35 region to the determined or predicted+1 TSS. The -35 element and the -10 element are shown in bold and the spacer length is indicated. A spacer length of 16 bp or a G at position -14 is shown in red. (**B**) Levels of RNA (average±standard deviation) from various promoters relative to the level observed for P_{S17} with σ^{70} D445 (black) or σ^{70} V445 (grey). RNAP was reconstituted and the reactions were performed at 37° C using 0.02 pmol of P_{S17} and P_{ampC} and 0.05 pmol for all other DNAs. The number of independent transcriptions (n) were: P_{S17} (D445, $n = 28$; V445, $n = 39$), P_{RNAI} ($n = 6$), P_{betI} ($n = 6$), $P_{RS00415}$ ($n = 4$), P_{xseA} ($n = 3$), P_{ampC} ($n = 10$), P_{lptM} ($n = 3$), P_{sppA} ($n = 3$), P_{yciY} ($n = 3$), P_{csrD} ($n = 3$), P_{yobH} ($n = 4$). P-values were generated by a *t*-test for two independent means; NS, not significant with *P*-value >0.05; *, *P*-value <0.05; **, *P*-value <0.01; ***, *P*-value <0.001; ****, *P*-value <0.0001. The fold difference for the level of RNA from the tested promoter versus the level from P_{S17} is given below the graph. (**C**) Representative gel slices showing the labeled RNA products. Expected sizes of the RNAs from the various promoters are indicated, and the positions of the RNAs from the tested promoters are denoted with arrows.

several properties that are associated with the LF82 lifestyle. For example, changes in antibiotic resistance profiles have been observed among *E. coli* associated with CD (79), and LF82 is known to be Amp^R (63). Similarly, a dramatic change afforded by σ^{70} V445 is the reprogramming of the overall antibiotic resistance pattern (Table 4), including resistance to β-lactams from the elevated expression of *ampC*, and other changes in antibiotic resistance presumably from the increased expression of multi-drug resistance efflux pumps.

LF82 is also known to have increased biofilm formation relative to commensal *E. coli* (68), a property that may contribute to survival within macrophages (80). Accordingly, $MG1655^{D445V}$ showed increased biofilm relative to MG1655 and a mucoid phenotype at lower temperatures. Biofilm production and mucoidy involve multiple, common genes within several operons (69,81). The alternative σ factor σ^s is thought to regulate both processes (69), but the mechanism of σ^s action is unclear as σ^s positively or negatively affects biofilm formation, depending on the specific conditions (69,82). In addition, the regulation of σ^s is complex, involving multiple factors that influence transcription, translation, and proteolysis (70). While the *rpoS* RNA levels were similar in both the MG1655 and LF82 RNA-seq datasets, we found lower amounts of σ^s protein in $MG1655^{D445V}$ relative to MG1655, indicating that σ^{70} V445 affects the post-transcriptional regulation of σ^s. Dysregulation of σ^s could then lead to downstream effects on biofilm formation and the mucoid phenotype.

Studies have found a connection between an increase in methionine and CD (83), including elevated levels of methionine and methionine metabolites in fecal material from CD patients and improved colitis symptoms when DSS-treated mice were placed on a methionine-restricted diet (84). Our analyses revealed that the presence of σ^{70} 445V both increases the level of *metE*, whose product catalyzes the final step in methionine biosynthesis in the absence of B12 [reviewed in (67)], and improves the ability of MG1655 to grow in the absence of methionine. Thus, with σ^{70} V445 the ability of *E. coli* to respond to conditions of methionine limitation is altered.

Flagellar assembly is vital to the LF82 lifestyle since LF82 uses type 1 pili, encoded by the *fim* genes, to adhere to the CEACAM 6 receptor on the intestinal epithelial cell surface (3,85). Thus, deletion of *flhDC* or *fliA*, which are needed for the expression of the *fim* operon, results in a decrease in adhesion and invasion (86), and deletion of *fliC*, encoding flagellin, renders LF82 unable to adhere to or invade Hep-2 epithelial cells (87). However, while motility is an asset in host-to-host transmission in mice (14), the ability to regulate flagella formation during chronic colonization is also important (85), since flagella are targeted by the host immune system.

A

	-35	spacer	-10	+1	Fold DIfference
P_{S17}	**TTGCAT**	CGCACAAAAGAATCCTC (17 bp)	**TACTAT**	TTGCACA	1.1X
P_{ampC}	**TTGACA**	CACTGATTGGTGTCGT (16 bp)	**TACAAT**	CTAACGCA	1.9X
$P_{ampC,S17}$	**TTGACA**	CGCACAAAAGAATCCTC (17 bp)	**TACAAT**	CTAACGCA	1.1X
P_{-14T}	**TTGACA**	CACTGATTGGTGTCTT (16 bp)	**TACAAT**	CTAACGCA	1.9X
P_{-14A}	**TTGACA**	CACTGATTGGTGTCAT (16 bp)	**TACAAT**	CTAACGCA	1.5X
P_{-14C}	**TTGACA**	CACTGATTGGTGTCCT (16 bp)	**TACAAT**	CTAACGCA	1.8X
P_{+14T}	**TTGACA**	CACTGATTGGTGTCGTT (17 bp)	**TACAAT**	CTAACGCA	1.0X
P_{+18A}	**TTGACA**	CACTGATTGGTAGTCGT (17 bp)	**TACAAT**	CTAACGCA	1.2X
P_{+16G}	**TTGACA**	CACTGATTGGTGTGCGT (17 bp)	**TACAAT**	CTAACGCA	1.1X
$P_{+18A,-14T}$	**TTGACA**	CACTGATTGGTAGTCTT (17 bp)	**TACAAT**	CTAACGCA	1.3X
$P_{+16G,-14T}$	**TTGACA**	CACTGATTGGTGTGCTT (17 bp)	**TACAAT**	CTAACGCA	1.0X

B

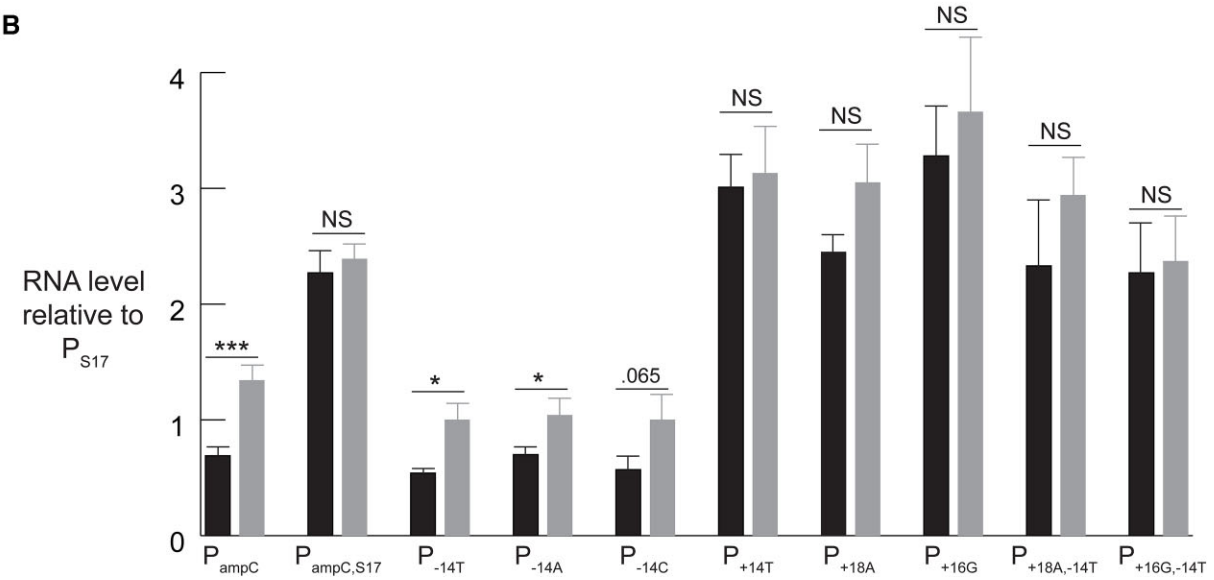


Figure 5. Increasing the spacer length of P_{ampC} from 16 to 17 bp eliminates the preference by σ^{70} V445-RNAP. (**A**) Sequences of P_{ampC} mutant promoters from the -35 region to the+1 TSS. The −35 element, the −10 element, and a G at −14 are shown in black bold and the spacer length is indicated. Sequences or spacer length that deviate from that of P_{ampC} are shown in red. The fold difference in the relative level of RNA (level compared to that from P_{S17}) from each promoter for σ^{70} V445-RNAP *vs.* σ^{70} D445-RNAP is indicated. (**B**) Levels of RNA (average±standard deviation) from the various promoters relative to the level observed for P_{S17} with σ^{70} D445-RNAP (black) or σ^{70} V445-RNAP (grey). RNAP was reconstituted and the reactions were performed at 37° C using 0.02 pmol of P_{S17} and 0.05 pmol for all other DNAs. The number of independent transcriptions (n) were P_{ampC} ($n = 4$), $P_{ampc,S17}$ ($n = 4$), P_{-14T} ($n = 3$), P_{-14A} ($n = 3$), P_{-14C} ($n = 3$), P_{+14T} ($n = 4$), P_{+18A} ($n = 3$), P_{+16G} ($n = 4$), $P_{+18A,-14T}$ ($n = 3$), $P_{+16G,-14T}$ ($n = 4$). P-values were generated by a *t*-test for two independent means; NS, not significant with *P*-value >0.05; *, *P*-value <0.05; ***, *P*-value <0.001; the *P*-value for P_{-15C} is given (0.065).

The alteration in the expression of multiple flagella and motility genes at 23° C and of *flhA*, which is required for flagellin export, at 37° C suggest that σ^{70} V445 may provide a way to regulate the levels of the highly immunogenic flagellar antigens.

The increased expression of three other genes, which are observed in both the MG1655 and LF82 backgrounds (*shiA*, *aroP*, and *yobH*), may also aid LF82 in its lifestyle. ShiA and AroP are involved in the transport of aromatic intermediates or tryptophan, respectively. Tryptophan levels are low in CD patients (88), suggesting a possible need for increased tryptophan import into the LF82 cell. YobH is an uncharacterized protein in *E. coli* but is required for *Salmonella entrica* to invade macrophages (61). Macrophage invasion is a property that distinguishes LF82 and other AIEC from commensal and laboratory *E. coli*.

Finally, our analyses also indicate that an even larger number of genes and 2 pathways, flagella/motility and ATP syn-

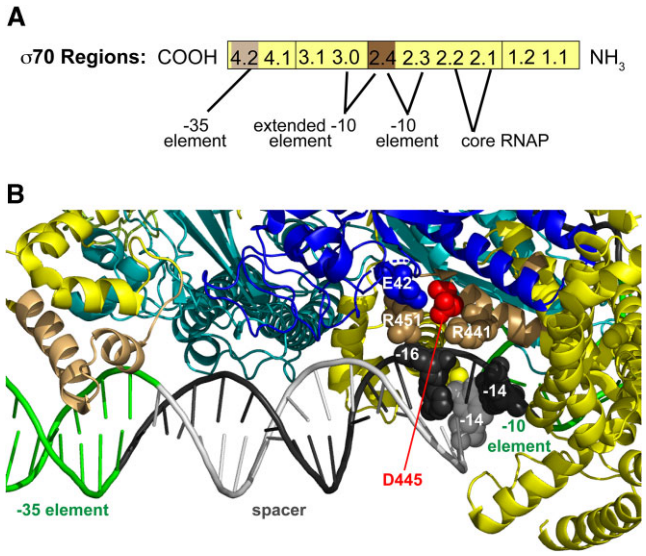


Figure 6. Regions of σ^{70} and position of D445 within the σ^{70}-RNAP structure. (**A**) Schematic of σ^{70} from the C-terminus (COOH) to the N-terminus (NH_3), showing the positions of σ^{70} regions (R). Residues within R4.2 interact with the -35 element, within R3.0/R2.4 interact with the extended -10 element, within R2.4/R2.3 interact with the -10 element, and within R2.2/R2.1 interact with core RNAP. R4.2 and R2.4 are indicated as light and dark brown, respectively. (**B**) Close up of the cryo-EM structure of *E. coli* RNAP with promoter DNA having a 17 bp spacer and showing promoter positions from the −35 element to the −10 element [PDB 6CA0 (19)]. Top (nontemplate) strand of the DNA is in black, bottom (template) strand is in grey except for the −35 and −10 elements, which are in green. Base positions at -14 (nontemplate black, template grey) and −16 (nontemplate black) are shown as spheres. σ^{70} colors correspond to those shown in (A) except for D445, which is shown as a red sphere; β′ is in cobalt blue and β is in cyan. The β′ residue E42 and the σ^{70} residues R441 and R451 are shown as spheres. (The RNAP subunits α_1, α_2, and ω cannot be seen in this closeup.)

thesis, are affected by the σ^{70} D445V substitution in MG1655 at 23°C. These results predict that σ^{70} V445 imparts additional changes that may aid the pathobiont outside the gut.

σ^{70} V445 alters the interaction of RNAP with specific promoters

Although σ^{70} is a highly conserved protein throughout eubacteria, certain regions (R) of σ^{70} that interact directly with core RNAP or with DNA promoter elements are nearly identical [reviewed in (18)]. These include residues within R2.1 and R2.2, which interact with the core polymerase; within R2.3 and R2.4, where specific residues contact the −10 sequence motif; and within R4.2, which interacts with the −35 element (Figure 6A). The -10 motif is typically located from −12 to −7. In addition, an upstream 'extended −10 motif' ($^{-15}$TG^{-14}), when present, can provide an additional contact with highly conserved residues in R2.4 and in R3.0. In particular, R2.4 residue R441 directly contacts the nontemplate (top) strand at position -14, with G being the consensus base (Figure 6B) (19,90). Finally, the distance between the −10 and −35 elements is crucial, with a spacer of 17 bp being optimal, but either 1 bp shorter (16 bp spacer) or 1 bp longer (18 bp) is still acceptable. This distance is imposed by the distance between σ^{70} R2.4 and σ^{70} R4.2, when they are present within RNAP.

Inspection of the cryo-EM structure of *E. coli* RNAP with promoter DNA [PDB 6CA0, (19)] and an x-ray crystallographic structure of *Thermus aquaticus* RNAP with promoter DNA [PDB 4XLN, (90)] reveals that D445 (red sphere in Figure 6B) is located within R2.4, but lies >15 Å from any bases within the −10 or extended −10 elements, a distance too far for the amino acid to directly interact with the DNA. D445 is, however, close enough (<5 Å) to interact with σ^{70} residues R441 and R451, which in the crystal structure (90) contact the nontemplate position −14 G and the nontemplate position -16, respectively. In addition, other biochemical work has suggested that R451 may affect the activity of certain promoters based on the nontemplate base at position −18 (91). D445 is also <5 Å from E42 within the β' subunit of RNAP (blue sphere in Figure 6B), which lies within the β' zipper, a region shown to interact with the promoter spacer region (90,92).

A crucial step in transcription initiation is the entry of the promoter DNA into the RNAP catalytic pocket where the double-stranded DNA from −11 to ∼+3 must unwind to form the transcription bubble. Since σ residues R441 and R451 can contact both D445 and the spacer DNA, we hypothesize that during this entry the negatively charged DNA and D445 compete for these positively-charged residues. As the sequence of the spacer sets the trajectory of the DNA entering the pocket (89,91,93), the spacer sequence can mediate this competition. In LF82 the rare nonpolar valine substitution at D445 eliminates this competition, which means that certain less favored trajectories are now more acceptable. Given that most of the preferred promoters had the 16 bp spacer/−14 G:C combination, it seems that the shorter spacer together with a strong contact at −14, provides a particular combination that is improved when using σ^{70} V445-RNAP. However, other combinations are also beneficial, as seen with P_{csrD} and P_{sppA}, and at P_{ampC}, the specific bp at −14 is not the deciding factor.

Our work adds to several studies indicating that the spacer sequence has a substantial impact on determining the activity of a promoter. Early work reported correlations between spacer sequence and activity for P_{lac} (94) and P1 of *gapA* (95). More recent work has shown that RNAP-spacer contacts are essential for initiation (96) and that D451 together with AT-richness between −17 and −23 can substantially increase promoter activity (97). Interestingly, having AT-richness from −17 to −23 does not correlate with the increased expression from the promoters with σ^{70} V445-RNAP (Figure 4). Nonetheless, in all of these cases, it is likely that the spacer sequence is affecting how the −10 element enters the transcription pocket and consequently, the process of initiation.

In conclusion, the discovery of a unique variant of the housekeeping σ^{70} factor of LF82 that imparts phenotypes consistent with the LF82 lifestyle suggests that the *rpoD* D445V allele aids in the emergence of a pathobiont strain that is uniquely suited to its role in the gut environment. These results suggest that the highly conserved housekeeping genes represent a target for genetic adaptation, leading to the acquisition of multiple phenotypes via a single base pair change. This may represent an under-studied mechanism for the emergence of new strain variants in nature.

Data availability

Genomic sequencing data for MG1655 and MG1655^{D445V} are deposited with NCBI BioSample (Accession numbers:

SAMN32661817-SAMN32661824). RNA-seq analyses of MG1655^{D445V} *vs*. MG1655 grown at 37°C and at 23°C and of LF82 *vs*. LF82^{V445D} grown in culture at 37°C are deposited in the NCBI database (GEO numbers GSE222248 and GSE253924, respectively). All other data are available in the main text or the supplementary materials.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We thank Michael Cashel for providing *E. coli* MG1655, Jeffers Nguyen for providing the plasmid pJN1, Nicholas Backes for performing genome editing to generate isogenic strains, Michael Margala for coding help, the NIDDK genomics core for RNA sequencing, the NCI CCR sequencing core (Caroline Fromont, Bao Tran, Xiongfong Chen, Yongmei Zhao, and Sulbha Choudhari) for whole genome sequencing, Errin Frahm and Robert Herbert of the NIDDK Computer Technology Branch for assistance with bioinformatic analysis, Xinglin Jia for assistance with BLAST search analysis, Ethan Pham for MIC tests with LF82, Leslie Knipling for generating plasmids for *in vitro* transcription, and Tom Schneider (NCI) and members of the Hinton and Phillips laboratory for helpful discussions and comments.

Funding

Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health (to M.A.M., A.C.-M., V.R., H.L., D.M.H.]; NIH [R01 GM099537 and R03 CA195305]; DOD-Defense Threat Reduction Agency (DTRA) [HDTRA12110015 to G.J.P.]. Funding for open access charge: Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health; NIH [R01 GM099537 and R03 CA195305]; DOD-Defense Threat Reduction Agency (DTRA) [HDTRA12110015].

Conflict of interest statement

None declared.

References

1. Torres,J., Mehandru,S., Colombel,J.F. and Peyrin-Biroulet,L. (2017) Crohn's disease. *Lancet*, **389**, 1741–1755.
2. Darfeuille-Michaud,A., Neut,C., Barnich,N., Lederman,E., Di Martino,P., Desreumaux,P., Gambiez,L., Joly,B., Cortot,A. and Colombel,J.F. (1998) Presence of adherent *Escherichia coli* strains in ileal mucosa of patients with Crohn's disease. *Gastroenterology*, **115**, 1405–1413.
3. Mirsepasi-Lauridsen,H.C., Vallance,B.A., Krogfelt,K.A. and Petersen,A.M. (2019) *Escherichia coli* pathobionts associated with inflammatory bowel disease. *Clin. Microbiol. Rev.*, **32**, e00060-18.
4. Shaler,C.R., Elhenawy,W. and Coombes,B.K. (2019) The unique lifestyle of Crohn's Disease-associated adherent-invasive *Escherichia coli*. *J. Mol. Biol.*, **431**, 2970–2981.
5. Renouf,M.J., Cho,Y.H. and McPhee,J.B. (2019) Emergent behavior of IBD-associated *Escherichia coli* during disease. *Inflamm. Bowel Dis.*, **25**, 33–44.
6. Glasser,A.L., Boudeau,J., Barnich,N., Perruchot,M.H., Colombel,J.F. and Darfeuille-Michaud,A. (2001) Adherent invasive *Escherichia coli* strains from patients with Crohn's disease survive and replicate within macrophages without inducing host cell death. *Infect. Immun.*, **69**, 5529–5537.
7. Bruder,E. and Espéli,O. (2022) *Escherichia coli* bacteria associated with Crohn's disease persist within phagolysosomes. *Curr. Opin. Microbiol.*, **70**, 102206.
8. Barnich,N., Carvalho,F.A., Glasser,A.L., Darcha,C., Jantscheff,P., Allez,M., Peeters,H., Bommelaer,G., Desreumaux,P., Colombel,J.F., *et al.* (2007) CEACAM6 acts as a receptor for adherent-invasive *E. coli*, supporting ileal mucosa colonization in Crohn disease. *J. Clin. Invest.*, **117**, 1566–1574.
9. Agus,A., Richard,D., Faïs,T., Vazeille,E., Chervy,M., Bonnin,V., Dalmasso,G., Denizot,J., Billard,E., Bonnet,R., *et al.* (2021) Propionate catabolism by CD-associated adherent-invasive *E. coli* counteracts its anti-inflammatory effect. *Gut Microbes*, **13**, 1–18.
10. Nash,J.H., Villegas,A., Kropinski,A.M., Aguilar-Valenzuela,R., Konczy,P., Mascarenhas,M., Ziebell,K., Torres,A.G., Karmali,M.A. and Coombes,B.K. (2010) Genome sequence of adherent-invasive *Escherichia coli* and comparative genomic analysis with other *E. coli* pathotypes. *Bmc Genomics [Electronic Resource]*, **11**, 667.
11. O'Brien,C.L., Bringer,M.A., Holt,K.E., Gordon,D.M., Dubois,A.L., Barnich,N., Darfeuille-Michaud,A. and Pavli,P. (2017) Comparative genomics of Crohn's disease-associated adherent-invasive *Escherichia coli*. *Gut*, **66**, 1382–1389.
12. Barrios-Villa,E., Martínez de la Peña,C.F., Lozano-Zaraín,P., Cevallos,M.A., Torres,C., Torres,A.G. and Rocha-Gracia,R.D.C. (2020) Comparative genomics of a subset of Adherent/Invasive *Escherichia coli* strains isolated from individuals without inflammatory bowel disease. *Genomics*, **112**, 1813–1820.
13. Martinez-Medina,M., Mora,A., Blanco,M., López,C., Pilar Alonso,M., Bonacorsi,S., Nicolas-Chanoine,M.-H., Darfeuille-Michaud,A., Garcia-Gil,J. and Blanco,J. (2009) Similarity and divergence among adherent-invasive *Escherichia coli* and extraintestinal pathogenic *E. coli* strains. *J. Clin. Microbiol.*, **47**, 3968–3979.
14. Elhenawy,W., Tsai,C.N. and Coombes,B.K. (2019) Host-specific adaptive diversification of Crohn's disease-associated adherent-invasive *Escherichia coli*. *Cell Host Microbe*, **25**, 301–312.
15. Proctor,A., Parvinroo,S., Richie,T., Jia,X., Lee,S.T.M., Karp,P.D., Paley,S., Kostic,A.D., Pierre,J.F., Wannemuehler,M.J., *et al.* (2022) Resources to facilitate use of the altered Schaedler flora (ASF) mouse model to study microbiome function. *mSystems*, **7**, e0029322.
16. Wymore Brand,M., Proctor,A.L., Hostetter,J.M., Zhou,N., Friedberg,I., Jergens,A.E., Phillips,G.J. and Wannemuehler,M.J. (2022) Vertical transmission of attaching and invasive *E. coli* from the dam to neonatal mice predisposes to more severe colitis following exposure to a colitic insult later in life. *PLoS One*, **17**, e0266005.
17. Small,C.L., Reid-Yu,S.A., McPhee,J.B. and Coombes,B.K. (2013) Persistent infection with Crohn's disease-associated adherent-invasive *Escherichia coli* leads to chronic inflammation and intestinal fibrosis. *Nat. Commun.*, **4**, 1957.
18. Decker,K.B. and Hinton,D.M. (2013) Transcription regulation at the core: similarities among bacterial, archaeal, and eukaryotic RNA polymerases. *Annu. Rev. Microbiol.*, **67**, 113–139.
19. Narayanan,A., Vago,F.S., Li,K., Qayyum,M.Z., Yernool,D., Jiang,W. and Murakami,K.S. (2018) Cryo-EM structure of *Escherichia coli* sigma(70) RNA polymerase and promoter DNA complex revealed a role of sigma non-conserved region during the open complex formation. *J. Biol. Chem.*, **293**, 7367–7375.
20. Boudeau,J., Glasser,A.L., Masseret,E., Joly,B. and Darfeuille-Michaud,A. (1999) Invasive ability of an *Escherichia coli* strain isolated from the ileal mucosa of a patient with Crohn's disease. *Infect. Immun.*, **67**, 4499–4509.
21. Studier,F.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.*, **185**, 60–89.
22. Kelly,P., Backes,N., Mohler,K., Buser,C., Kavoor,A., Rinehart,J., Phillips,G. and Ibba,M. (2019) Alanyl-tRNA synthetase quality

control prevents global dysregulation of the *Escherichia coli* proteome. *mBio*, **10**, e02921-19.
23. Ferrieres,L., Hemery,G., Nham,T., Guerout,A.M., Mazel,D., Beloin,C. and Ghigo,J.M. (2010) Silent mischief: bacteriophage mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *J. Bacteriol.*, **192**, 6418–6427.
24. Kim,J., Webb,A.M., Kershner,J.P., Blaskowski,S. and Copley,S.D. (2014) A versatile and highly efficient method for scarless genome editing in *Escherichia coli* and *Salmonella enterica*. *BMC Biotechnol.*, **14**, 84.
25. Cooper,H.S., Murthy,S.N., Shah,R.S. and Sedergran,D.J. (1993) Clinicopathologic study of dextran sulfate sodium experimental murine colitis. *Lab. Invest.*, **69**, 238–249.
26. Mahler,M., Bristol,I.J., Leiter,E.H., Workman,A.E., Birkenmeier,E.H., Elson,C.O. and Sundberg,J.P. (1998) Differential susceptibility of inbred mouse strains to dextran sulfate sodium-induced colitis. *Am. J. Physiol.*, **274**, G544–G551.
27. Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
28. Pruitt,K., Brown,G., Tatusova,T. and Maglott,D. (2002) In: *The NCBI Handbook*. National Center for Biotechnology Information (US), Bethesda, MD, pp. 1–22.
29. Van der Auwera,G.A., Carneiro,M.O., Hartl,C., Poplin,R., Del Angel,G., Levy-Moonshine,A., Jordan,T., Shakir,K., Roazen,D., Thibault,J., *et al.* (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics*, **43**, 11.10.1–11.10.33.
30. DePristo,M.A., Banks,E., Poplin,R., Garimella,K.V., Maguire,J.R., Hartl,C., Philippakis,A.A., Angel,G., Rivas,M.A., Hanna,M., *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.*, **43**, 491–498.
31. McKenna,A., Hanna,M., Banks,E., Sivachenko,A., Cibulskis,K., Kernytsky,A., Garimella,K., Altshuler,D., Gabriel,S., Daly,M., *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, **20**, 1297–1303.
32. Lassmann,T., Hayashizaki,Y. and Daub,C.O. (2011) SAMStat: monitoring biases in next generation sequencing data. *Bioinformatics*, **27**, 130–131.
33. Kearse,M., Moir,R., Wilson,A., Stones-Havas,S., Cheung,M., Sturrock,S., Buxton,S., Cooper,A., Markowitz,S., Duran,C., *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.
34. Madden,T. (2002) In: McEntyre,J. and Ostell,J. (eds.) *The NCBI Handbook*. National Center for Biotechnology Information (US), MD.
35. Bonocora,R.P., Decker,P.K., Glass,S., Knipling,L. and Hinton,D.M. (2011) Bacteriophage T4 MotA activator and the beta-flap tip of RNA polymerase target the same set of sigma70 carboxyl-terminal residues. *J. Biol. Chem.*, **286**, 39290–39296.
36. Elliott,T. and Geiduschek,E.P. (1984) Defining a bacteriophage T4 late promoter: absence of a "-35" region. *Cell*, **36**, 211–219.
37. Morita,M. and Oka,A. (1979) The structure of a transcriptional unit on colicin E1 plasmid. *Eur. J. Biochem.*, **97**, 435–443.
38. Gerber,J.S. and Hinton,D.M. (1996) An N-terminal mutation in the bacteriophage T4 *motA* gene yields a protein that binds DNA but is defective for activation of transcription. *J. Bacteriol.*, **178**, 6133–6139.
39. Baxter,K., Lee,J., Minakhin,L., Severinov,K. and Hinton,D.M. (2006) Mutational analysis of sigma(70) region 4 needed for appropriation by the bacteriophage T4 transcription factors AsiA and MotA. *J. Mol. Biol.*, **363**, 931–944.
40. Hinton,D.M. (1989) Transcript analyses of the uvsX-40-41 region of bacteriophage T4. Changes in the RNA as infection proceeds. *J. Biol. Chem.*, **264**, 14432–14439.
41. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
42. Liao,Y., Smyth,G.K. and Shi,W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.
43. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
44. Stephens,M. (2017) False discovery rates: a new deal. *Biostatistics*, **18**, 275–294.
45. Virtanen,P., Gommers,R., Oliphant,T.E., Haberland,M., Reddy,T., Cournapeau,D., Burovski,E., Peterson,P., Weckesser,W., Bright,J., *et al.* (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods*, **17**, 261–272.
46. Sanchez-Vazquez,P., Dewey,C.N., Kitten,N., Ross,W. and Gourse,R.L. (2019) Genome-wide effects on *Escherichia coli* transcription from ppGpp binding to its two sites on RNA polymerase. *Proc. Natl. Acad. Sci. U S A*, **116**, 8310–8319.
47. Keseler,I.M., Gama-Castro,S., Mackie,A., Billington,R., Bonavides-Martinez,C., Caspi,R., Kothari,A., Krummenacker,M., Midford,P.E., Muniz-Rascado,L., *et al.* (2021) The EcoCyc Database in 2021. *Front. Microbiol.*, **12**, 711077.
48. Keseler,I.M., Mackie,A., Santos-Zavaleta,A., Billington,R., Bonavides-Martinez,C., Caspi,R., Fulcher,C., Gama-Castro,S., Kothari,A., Krummenacker,M., *et al.* (2017) The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res.*, **45**, D543–D550.
49. McKinney,W. (2010) Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*. **445**, 51–61.
50. Hunter,J.D. (2007) Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.*, **9**, 90–95.
51. Polzin,S., Huber,C., Eylert,E., Elsenhans,I., Eisenreich,W. and Schmidt,H. (2013) Growth media simulating ileal and colonic environments affect the intracellular proteome and carbon fluxes of enterohemorrhagic *Escherichia coli* O157:H7 strain EDL933. *Appl. Environ. Microbiol.*, **79**, 3703–3715.
52. Carvalho,F.A., Barnich,N., Sivignon,A., Darcha,C., Chan,C.H., Stanners,C.P. and Darfeuille-Michaud,A. (2009) Crohn's disease adherent-invasive *Escherichia coli* colonize and induce strong gut inflammation in transgenic mice expressing human CEACAM. *J. Exp. Med.*, **206**, 2179–2189.
53. Schaedler,R.W., Dubos,R. and Costello,R. (1965) The development of the bacterial flora in the gastrointestinal tract of mice. *J. Exp. Med.*, **122**, 59–66.
54. Paget,M.S. (2015) Bacterial sigma factors and anti-sigma factors: structure, function and distribution. *Biomolecules*, **5**, 1245–1265.
55. Desilets,M., Deng,X., Rao,C., Ensminger,A.W., Krause,D.O., Sherman,P.M. and Gray-Owen,S.D. (2016) Genome-based definition of an inflammatory bowel disease-associated adherent-invasive *Escherichia coli* pathovar. *Inflamm. Bowel Dis.*, **22**, 1–12.
56. Miquel,S., Peyretaillade,E., Claret,L., de Vallee,A., Dossat,C., Vacherie,B., Zineb el,H., Segurens,B., Barbe,V., Sauvanet,P., *et al.* (2010) Complete genome sequence of Crohn's disease-associated adherent-invasive *E. coli* strain LF82. *PLoS One*, **5**, e12714.
57. Zhang,Y., Rowehl,L., Krumsiek,J.M., Orner,E.P., Shaikh,N., Tarr,P.I., Sodergren,E., Weinstock,G.M., Boedeker,E.C., Xiong,X., *et al.* (2015) Identification of candidate adherent-invasive *E. coli* signature transcripts by genomic/transcriptomic analysis. *PLoS One*, **10**, e0130902.
58. Fang,X., Monk,J.M., Nurk,S., Akseshina,M., Zhu,Q., Gemmell,C., Gianetto-Hill,C., Leung,N., Szubin,R., Sanders,J., *et al.* (2018) Metagenomics-based, strain-level analysis of *Escherichia coli* from a time-series of microbiome samples from a Crohn's Disease patient. *Front. Microbiol.*, **9**, 2559.
59. Ohkusa,T., Okayasu,I., Ogihara,T., Morita,K., Ogawa,M. and Sato,N. (2003) Induction of experimental ulcerative colitis by

Fusobacterium varium isolated from colonic mucosa of patients with ulcerative colitis. *Gut*, **52**, 79–83.
60. Battesti,A., Majdalani,N. and Gottesman,S. (2011) The RpoS-mediated general stress response in *Escherichia coli*. *Annu. Rev. Microbiol.*, **65**, 189–213.
61. Banda,M.M., Manzo,R. and Bustamante,V.H. (2019) HilD induces expression of a novel *Salmonella typhimurium* invasion factor, YobH, through a regulatory cascade involving SprB. *Sci. Rep.*, **9**, 12725.
62. Riley,M., Abe,T., Arnaud,M.B., Berlyn,M.K., Blattner,F.R., Chaudhuri,R.R., Glasner,J.D., Horiuchi,T., Keseler,I.M., Kosuge,T., *et al.* (2006) *Escherichia coli* K-12: a cooperatively developed annotation snapshot–2005. *Nucleic Acids Res.*, **34**, 1–9.
63. Martinez-Medina,M., Strozzi,F., Castillo,B.R.D., Serrano-Morillas,N., Bustins,N.F. and Martínez-Martínez,L. (2020) Antimicrobial resistance profiles of adherent invasive *Escherichia coli* show increased resistance to β-lactams. *Antibiotics (Basel)*, **9**, 251.
64. (2020) *The European Committee on Antibicrobial Suspeptibility Testing*. Vol. version 10.0.
65. Nishino,K. and Yamaguchi,A. (2001) Analysis of a complete library of putative drug transporter genes in *Escherichia coli*. *J. Bacteriol.*, **183**, 5803–5812.
66. Long,F., Rouquette-Loughlin,C., Shafer,W.M. and Yu,E.W. (2008) Functional cloning and characterization of the multidrug efflux pumps NorM from *Neisseria gonorrhoeae* and YdhE from *Escherichia coli*. *Antimicrob. Agents Chemother.*, **52**, 3052–3060.
67. Hondorp,E.R. and Matthews,R.G. (2006) Methionine. *EcoSal Plus*, **2**, https://doi.org/10.1128/ecosalplus.3.6.1.7.
68. Martinez-Medina,M., Naves,P., Blanco,P., Aldeguer,X., Blanco,J.E., Blanco,M., Ponte,C., Soriano,F., Darfeuille-Michaud,A. and Garcia-Gil,L.J. (2009) Biofilm formation as a novel phenotypic feature of adherent-invasive *Escherichia coli (*AIEC). *BMC Microbiol.*, **9**, 202.
69. Beloin,C., Roux,A. and Ghigo,J.M. (2008) *Escherichia coli* biofilms. *Curr. Top. Microbiol. Immunol.*, **322**, 249–289.
70. Gottesman,S. (2019) Trouble is coming: signaling pathways that regulate general stress responses in bacteria. *J. Biol. Chem.*, **294**, 11685–11700.
71. Wong,G.T., Bonocora,R.P., Schep,A.N., Beeler,S.M., Lee Fong,A.J., Shull,L.M., Batachari,L.E., Dillon,M., Evans,C., Becker,C.J., *et al.* (2017) Genome-wide transcriptional response to varying RpoS levels in *Escherichia coli* K-12. *J. Bacteriol.*, **199**, e00755-16.
72. Mitchell,J.E., Zheng,D., Busby,S.J. and Minchin,S.D. (2003) Identification and analysis of 'extended -10' promoters in *Escherichia coli*. *Nucleic Acids Res.*, **31**, 4689–4695.
73. Shultzaberger,R.K., Chen,Z., Lewis,K.A. and Schneider,T.D. (2007) Anatomy of *Escherichia coli* sigma70 promoters. *Nucleic Acids Res.*, **35**, 771–788.
74. Fabich,A.J., Leatham,M.P., Grissom,J.E., Wiley,G., Lai,H., Najar,F., Roe,B.A., Cohen,P.S. and Conway,T. (2011) Genotype and phenotypes of an intestine-adapted *Escherichia coli* K-12 mutant selected by animal passage for superior colonization. *Infect. Immun.*, **79**, 2430–2439.
75. Chen,L., Wang,D., Garmaeva,S., Kurilshikov,A., Vich Vila,A., Gacesa,R., Sinha,T., Segal,E., Weersma,R.K., Wijmenga,C., *et al.* (2021) The long-term genetic stability and individual specificity of the human gut microbiome. *Cell*, **184**, 2302–2315.
76. Greenblum,S., Carr,R. and Borenstein,E. (2015) Extensive strain-level copy-number variation across human gut microbiome species. *Cell*, **160**, 583–594.
77. Didelot,X., Walker,A.S., Peto,T.E., Crook,D.W. and Wilson,D.J. (2016) Within-host evolution of bacterial pathogens. *Nat. Rev. Microbiol.*, **14**, 150–162.
78. Tyakht,A.V., Manolov,A.I., Kanygina,A.V., Ischenko,D.S., Kovarsky,B.A., Popenko,A.S., Pavlenko,A.V., Elizarova,A.V., Rakitina,D.V., Baikova,J.P., *et al.* (2018) Genetic diversity of *Escherichia coli* in gut microbiota of patients with Crohn's disease discovered using metagenomic and genomic analyses. *Bmc Genomics [Electronic Resource]*, **19**, 968.
79. Dogan,B., Scherl,E., Bosworth,B., Yantiss,R., Altier,C., McDonough,P.L., Jiang,Z.D., Dupont,H.L., Garneau,P., Harel,J., *et al.* (2013) Multidrug resistance is common in *Escherichia coli* associated with ileal Crohn's disease. *Inflamm. Bowel Dis.*, **19**, 141–150.
80. Prudent,V., Demarre,G., Vazeille,E., Wery,M., Quenech'Du,N., Ravet,A., Dauverd-Girault,J., van Dijk,E., Bringer,M.A., Descrimes,M., *et al.* (2021) The Crohn's disease-related bacterial strain LF82 assembles biofilm-like communities to protect itself from phagolysosomal attack. *Commun. Biol.*, **4**, 627.
81. Ionescu,M. and Belkin,S. (2009) Overproduction of exopolysaccharides by an *Escherichia coli* K-12 *rpoS* mutant in response to osmotic stress. *Appl. Environ. Microb.*, **75**, 483–492.
82. Mika,F. and Hengge,R. (2013) Small regulatory RNAs in the control of motility and biofilm formation in *E. coli* and *Salmonella*. *Int. J. Mol. Sci.*, **14**, 4560–4579.
83. Walker,A. and Schmitt-Kopplin,P. (2021) The role of fecal sulfur metabolome in inflammatory bowel diseases. *Int. J. Med. Microbiol.*, **311**, 151513.
84. Liu,G., Yu,L., Fang,J., Hu,C.A., Yin,J., Ni,H., Ren,W., Duraipandiyan,V., Chen,S., Al-Dhabi,N.A., *et al.* (2017) Methionine restriction on oxidative stress and immune response in dss-induced colitis mice. *Oncotarget*, **8**, 44511–44520.
85. Sevrin,G., Massier,S., Chassaing,B., Agus,A., Delmas,J., Denizot,J., Billard,E. and Barnich,N. (2020) Adaptation of adherent-invasive *E. coli* to gut environment: impact on flagellum expression and bacterial colonization ability. *Gut Microbes*, **11**, 364–380.
86. Claret,L., Miquel,S., Vieille,N., Ryjenkov,D.A., Gomelsky,M. and Darfeuille-Michaud,A. (2007) The flagellar sigma factor FliA regulates adhesion and invasion of Crohn disease-associated *Escherichia coli* via a cyclic dimeric GMP-dependent pathway. *J. Biol. Chem.*, **282**, 33275–33283.
87. Barnich,N., Boudeau,J., Claret,L. and Darfeuille-Michaud,A. (2003) Regulatory and functional co-operation of flagella and type 1 pili in adhesive and invasive abilities of AIEC strain LF82 isolated from a patient with Crohn's disease. *Mol. Microbiol.*, **48**, 781–794.
88. Nikolaus,S., Schulte,B., Al-Massad,N., Thieme,F., Schulte,D.M., Bethge,J., Rehman,A., Tran,F., Aden,K., Hasler,R., *et al.* (2017) Increased tryptophan metabolism is associated with activity of inflammatory bowel diseases. *Gastroenterology*, **153**, 1504–1516.
89. Hook-Barnard,I.G. and Hinton,D.M. (2009) The promoter spacer influences transcription initiation via sigma70 region 1.1 of *Escherichia coli* RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 737–742.
90. Bae,B., Feklistov,A., Lass-Napiorkowska,A., Landick,R. and Darst,S.A. (2015) Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *eLife*, **4**, e08504.
91. Singh,S.S., Typas,A., Hengge,R. and Grainger,D.C. (2011) *Escherichia coli* sigma70 senses sequence and conformation of the promoter spacer region. *Nucleic Acids Res.*, **39**, 5109–5118.
92. Yuzenkova,Y., Tadigotla,V.R., Severinov,K. and Zenkin,N. (2011) A new basal promoter element recognized by RNA polymerase core enzyme. *EMBO J.*, **30**, 3766–3775.
93. Sharma,S., Pant,P., Arya,R., Jayaram,B. and Kumar Das,H. (2023) Bases immediate upstream of the TATAAT box of the sigma70 promoter of *Escherichia coli* significantly influence the activity of a model promoter by altering the bending angle of DNA. *Gene*, **851**, 146968.
94. Liu,M., Tolstorukov,M., Zhurkin,V., Garges,S. and Adhya,S. (2004) A mutant spacer sequence between -35 and -10 elements makes the P_{lac} promoter hyperactive and cAMP receptor protein-independent. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 6911–6916.
95. Thouvenot,B., Charpentier,B. and Branlant,C. (2004) The strong efficiency of the *Escherichia coli gapA* P1 promoter depends on a

complex combination of functional determinants. *Biochem. J.*, **383**, 371–382.
96. Meng,C.A., Fazal,F.M. and Block,S.M. (2017) Real-time observation of polymerase-promoter contact remodeling during transcription initiation. *Nat. Commun.*, **8**, 1178.
97. Warman,E.A., Singh,S.S., Gubieda,A.G. and Grainger,D.C. (2020) A non-canonical promoter element drives spurious transcription of horizontally acquired bacterial genes. *Nucleic Acids Res.*, **48**, 4891–4901.

Received: June 12, 2024. **Revised:** August 8, 2024. **Editorial Decision:** August 10, 2024. **Accepted:** August 28, 2024
Published by Oxford University Press on behalf of Nucleic Acids Research 2024.