Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Regulatory Evolution of TEM β-lactamases through Large-Scale Genomic Analysis Reveals Promoter Diversity and its Impact on β-lactam Resistance in Escherichia coli

Srimathy Ramachandran
SASTRA University
Snekha Arumugam Chandran
SASTRA University
Suma Mohan S.
sumamohan@scbt.sastra.edu

SASTRA University

Article

Keywords:

Posted Date: September 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10980732/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Regulatory Evolution of TEM β-lactamases through Large-Scale Genomic Analysis Reveals Promoter Diversity and its Impact on β-lactam Resistance in *Escherichia coli*

Srimathy Ramachandran[1], Snekha Arumugam Chandran[2], Suma Mohan S.[1]*

[1]*Computational Biology lab, Bioinformatics Center, School of Chemical and Biotechnology, SASTRA Deemed to be University, Thanjavur 613401, Tamil Nadu, India.*

[2]*Department of Bioinformatics, School of Chemical and Biotechnology, SASTRA Deemed to be University, Thanjavur 613401, Tamil Nadu, India.*

[*]Corresponding author: Suma Mohan S. : sumamohan@scbt.sastra.edu

Background: Antimicrobial resistance is a major global health threat, with β-lactamases representing the most prevalent mechanism of resistance to β-lactam antibiotics in Gram-negative bacteria. Among these, TEM β-lactamases (*bla*$_{\text{TEM}}$) are the earliest identified and most widely disseminated resistance determinants in *Escherichia coli*. Given the established influence of canonical promoter variation on *bla*$_{\text{TEM}}$ expression and β-lactam resistance, evolutionary diversification of the *bla*$_{\text{TEM}}$ upstream regulatory region can affect the resistance phenotype. However, the extent to which upstream variation is associated with resistance beyond changes in the coding sequence remains unclear.

Results: We analyzed 60,087 *E. coli* genomes from BV-BRC to characterize *bla*$_{\text{TEM}}$ upstream regulatory regions, focusing on -35 and -10 promoter motifs and their relationship with β-lactam resistance. P3 was the prevalent canonical promoter, while multiple non-canonical promoter types with distinct -10 and -35 motifs were identified. A causal-inference framework estimated promoter-associated differences in resistance probability while holding the underlying *bla*$_{\text{TEM}}$ variant profile constant. Average treatment effect estimates showed that several P3 transitions were associated with increased predicted resistance probabilities across multiple β-lactam antibiotics. Novel motifs from single-*bla*$_{\text{TEM}}$ isolates were further prioritized by occurrence frequency and evaluated structurally with σ^{70} DNA-recognition domains using molecular dynamics simulations. Notably, the -35 TGCAAC motif maintained consistent σ^{70} interactions, supporting its structural compatibility as a putative promoter motif potentially relevant to *bla*$_{\text{TEM}}$ regulation and resistance-associated phenotypes.

Conclusion: Overall, this study integrates large-scale genomics, causal inference, and structural analysis to investigate regulatory variation alongside coding-sequence changes, providing insight into the evolving mechanisms of antimicrobial resistance.

Background

Antimicrobial resistance (AMR) is a major global health challenge that compromises the effectiveness of antibiotics and substantially increases the burden of infectious diseases worldwide (World Health Organization, 2023). AMR primarily arises through the acquisition and dissemination of antibiotic resistance genes that enable bacteria to inactivate or neutralize antibiotics, modify or protect drug targets, reduce intracellular antibiotic concentrations through decreased permeability or active efflux, or circumvent drug action through metabolic bypass pathways (1). While the horizontal transfer of resistance genes is widely recognized as a major driver of AMR dissemination, the level at which these genes are expressed is equally important in determining the resistance phenotype (2). In bacteria, transcriptional regulation is primarily governed by promoter sequences that control when and to what extent genes are transcribed, thereby influencing bacterial adaptation under antibiotic selection pressure (3,4).

Among the diverse mechanisms of antimicrobial resistance, enzymatic inactivation of antibiotics by β-lactamases represents one of the most prevalent and clinically significant resistance strategies. β-lactamases hydrolyze the conserved β-lactam ring of β-lactam antibiotics, thereby preventing the antibiotics from binding to penicillin-binding proteins and inhibiting bacterial cell wall synthesis (5,6). Based on amino acid sequence homology, β-lactamases are classified into four molecular classes (A-D) according to the Ambler classification (7,8). Classes A, C, and D are serine β-lactamases that utilize a conserved active-site serine residue for catalysis, whereas Class B enzymes are metallo-β-lactamases that require zinc ions for hydrolysis (6). Among these, Ambler Class A β-lactamases include the TEM family, one of the earliest discovered and most extensively studied β-lactamase families and an important group of resistance determinants in Gram-negative bacteria (9–11). Given the widespread dissemination and continued diversification of *bla*$_{\text{TEM}}$ genes, this study focuses on the evolution of their upstream regulatory regions, with particular emphasis on promoter diversity and its potential contribution to β-lactam resistance. The prototype enzyme, *bla*$_{\text{TEM-1}}$, was first identified in the early 1960s from an *Escherichia coli* isolate obtained from a patient named Temoniera, from whom the designation "TEM" was derived (9). Initially, *bla*$_{\text{TEM-1}}$ conferred resistance primarily to narrow-spectrum penicillins, including ampicillin and amoxicillin, and early-generation cephalosporins through hydrolysis of the β-lactam ring (8,10,12). Continued antibiotic selection pressure has subsequently driven the diversification of the *bla*$_{\text{TEM}}$ family, resulting in more than 240 characterized *bla*$_{\text{TEM}}$ variants with diverse functional properties, including extended-spectrum β-lactamases capable of hydrolyzing third-generation cephalosporins and inhibitor-resistant variants with reduced susceptibility to β-lactamase inhibitors (8,13,14). In addition to coding sequence diversification, the successful dissemination of *bla*$_{\text{TEM}}$ has been facilitated by its frequent association with conjugative plasmids, particularly IncF plasmids in *E. coli*, as well as other plasmid incompatibility groups including IncI1, IncA/C2, IncL/M, IncN, IncX, and Col-type plasmids. Furthermore, *bla*$_{\text{TEM}}$ is commonly embedded within Tn3-family transposons and associated with insertion sequences such as IS26, enabling efficient horizontal transfer between plasmids and bacterial chromosomes (15–18).

Expression of *bla*$_{\text{TEM}}$ is primarily controlled by promoter sequences located upstream of the coding region. Transcription from these promoters is mediated by the housekeeping sigma factor σ^{70}, which recognizes the conserved -35 and -10 promoter elements and recruits RNA polymerase to initiate transcription(3,4). In *E. coli*, molecular characterization of *bla*$_{\text{TEM}}$-

associated promoter regions has identified four major naturally occurring promoter types - P3, Pa/Pb, P4, and P5 that differ by only one or a few nucleotide substitutions within the promoter sequence (19). Despite these relatively small sequence differences, these promoter variants exhibit distinct transcriptional activities. The ancestral P3 promoter confers lower basal expression, whereas promoter variants such as Pa/Pb, P4, and P5 can produce progressively higher levels of *bla*$_{TEM}$ expression and β-lactamase activity (19,20). Although several naturally occurring *bla*$_{TEM}$ promoters have been identified and experimentally characterized, their prevalence, evolutionary relationships, and sequence variation have not been systematically investigated using large-scale genomic datasets.

To systematically investigate the evolutionary diversity of the upstream regulatory regions of *bla*$_{TEM}$, we performed a large-scale genomic analysis of around 60,087 *E. coli* isolates obtained from the Bacterial and Viral Bioinformatics Resource Center (BV-BRC). We first characterized the prevalence, sequence diversity, and evolutionary relationships of canonical promoter types (P3, Pa/Pb, P4, and P5) and previously unreported promoter variants. We then examined insertion sequence (IS) elements within a 1,000-bp region flanking *bla*$_{TEM}$ to investigate the mobile genetic contexts associated with different promoter variants. Next, we applied a causal-inference framework to estimate promoter-associated differences in antimicrobial resistance phenotypes while considering a constant *bla*$_{TEM}$ variant profile. To further examine promoter variation in a less complex genetic background, the analysis was extended to isolates carrying a single *bla*$_{TEM}$ gene, enabling the prioritization of putative novel −35 and −10 promoter motifs. These selected motifs were subsequently subjected to structural analysis of σ^{70}-promoter interactions to investigate the binding possibility. Together, these analyses provide an integrated genomic, evolutionary, causal-inference, and structural framework for investigating the contribution of *bla*$_{TEM}$ promoter evolution to antimicrobial resistance in *E. coli*.

Results

Genomic distribution and diversity of *bla*$_{TEM}$ variants in global *E. coli* isolates

A large-scale global evolutionary analysis of the regulatory region of *E. coli bla*$_{TEM}$ genes were performed to investigate their potential relationship with antimicrobial resistance. A total of 60,087 *E. coli* isolates were collected from BV-BRC including 2,548 isolates containing explicitly annotated plasmid sequences. On performing screening for *bla*$_{TEM}$ β-lactamase genes, we found nearly 20,317 genome isolates carrying 21,367 chromosome-derived sequences and 965 genome isolates carrying 1,242 plasmid-derived *bla*$_{TEM}$ sequences (Fig. 1a). Overall, 78 distinct *bla*$_{TEM}$ variants were detected on chromosomes, and 24 on plasmids with *bla*$_{TEM-1}$ and *bla*$_{TEM-150}$ being the most prevalent variants in both genomic contexts. Based on the available metadata, *bla*$_{TEM}$ positive isolates ranged from the year 1951 to 2024 in chromosome and from 1974 to 2024 in plasmid, with the highest number recorded in 2019 in common. China and the United States contributed the largest and second-largest number of isolates, respectively. The number of *bla*$_{TEM}$ occurrences per isolate in chromosome ranged from 1 to 38, with observed counts of 1, 2, 3, 4, 5, 6, 7, 10, 12, 16, 19, 38 and in occurrences in plasmid ranged from 1 to 35 with the counts of 1,2,3,4,5,8,35. The metadata along with promoter motifs details for chromosome and plasmid reported in additional file (Additional file 1: Table S1a, S1b).

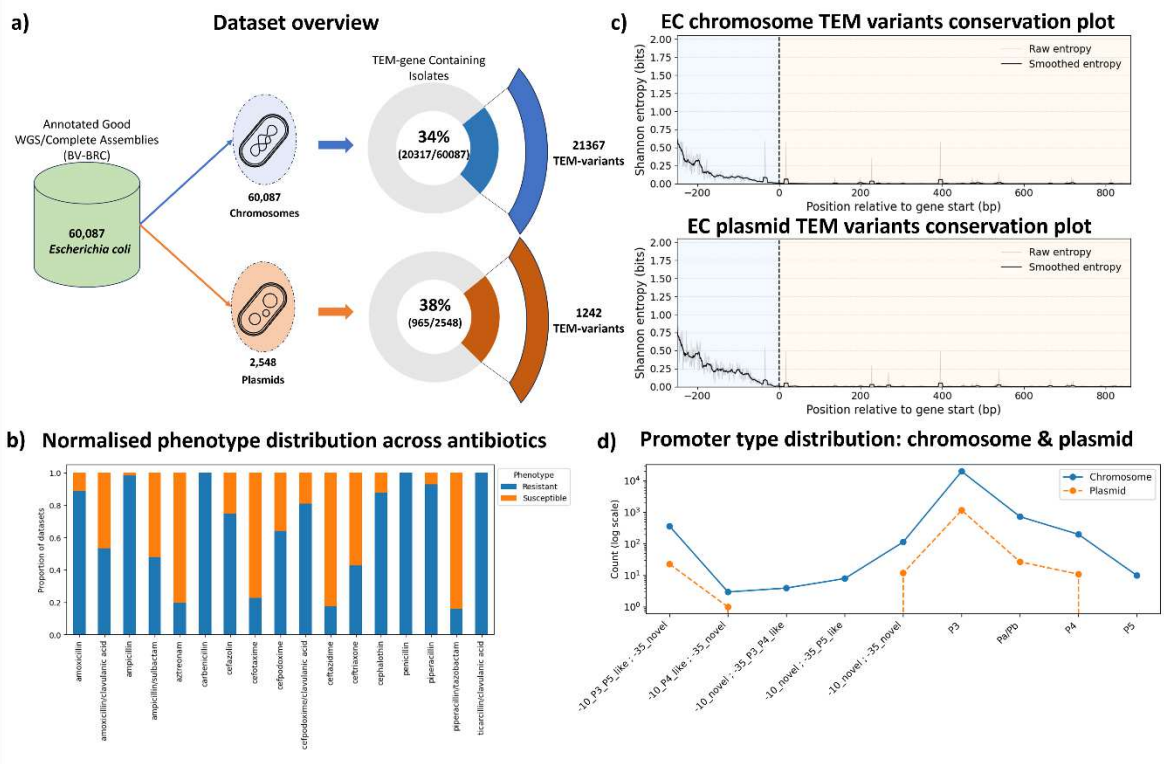


Fig. 1 Overview of the *bla*TEM dataset, antimicrobial susceptibility phenotypes, sequence conservation, and promoter distribution in *Escherichia coli*. **(a)** Overview of the analyzed dataset, showing the total number of chromosomal and plasmid sequences, the proportion of TEM-positive isolates, and the corresponding number of detected *bla*TEM genes. **(b)** Normalized distribution of resistant and susceptible phenotypes among TEM-positive *E. coli* isolates across β-lactam antibiotics. **(c)** Shannon entropy-based conservation profiles of *bla*TEM and its upstream regulatory region in chromosomal and plasmid sequences, with nucleotide positions shown relative to the gene start site. **(d)** Distribution of *bla*TEM promoter types among chromosomal and plasmid-associated sequences.

Antibiotic resistance phenotype distribution among *bla*TEM-positive isolates

The *bla*TEM-positive isolates were analyzed to investigate the antibiotic resistance phenotypes. The analysis included antibiotics and antibiotic-inhibitor combinations relevant to *bla*TEM-mediated beta-lactam resistance: amoxicillin, amoxicillin/clavulanic acid, ampicillin, ampicillin/sulbactam, aztreonam, carbenicillin, cefazolin, cefotaxime, cefpodoxime, cefpodoxime/clavulanic acid, ceftazidime, ceftriaxone, cephalothin, penicillin, piperacillin, piperacillin/tazobactam, ticarcillin/clavulanic acid. Isolates were classified into resistant and susceptible based on the available phenotypic data for each antibiotic. The normalized phenotype distribution showed that *bla*TEM genes were present in both resistant and susceptible isolates across all antibiotics examined (Fig. 1b). High proportions of resistance were observed for several penicillins and first-generation cephalosporins, including amoxicillin, ampicillin, carbenicillin, cephalothin, piperacillin, and ticarcillin/clavulanic acid. In contrast, aztreonam, ceftazidime, cefotaxime, and several β-lactam/β-lactamase inhibitor combinations showed relatively higher proportions of susceptible isolates despite the presence of *bla*TEM genes.

Evolutionary conservation and regulatory diversity across *bla*$_{TEM}$ upstream and coding regions

To investigate the evolutionary dynamics of *bla*$_{TEM}$ genes in the collected *E. coli* isolates, the upstream regulatory region and the coding regions were extracted to understand the sequence conservation and variability present in both chromosome and plasmid. Position-wise conservation analysis was performed on 21,114 chromosome-derived and on 1,242 plasmid-derived TEM gene, focussing on a minimum of 250 bp upstream of the gene start site and the complete coding region (Fig. 1c). In both genomic contexts, entropy-based conservation analysis revealed distinct differences between the upstream and the *bla*$_{TEM}$ coding region. The upstream region (∼-250 to -50 bp) exhibited relatively higher and more variable shannon entropy values, indicating increased sequence diversity. When comparing these regions between genomic contexts, shannon entropy of plasmid (~0.75 bits) shows higher variability in the upstream sequences when compared to chromosomes (~0.55 bits). Notably, entropy gradually decreased closer to the gene start site (∼-50 to 0 bp) suggesting increased conservation of proximal regulatory elements involved in transcription initiation. In contrast, the *bla*$_{TEM}$ coding region (0 to ∼850 bp) displayed consistently low entropy across the entire gene, indicating strong sequence conservation in both chromosomal and plasmid sequences. Only minor entropy peaks were observed at specific positions, likely corresponding to known variant-defining mutations, while the overall sequence remained highly conserved in both chromosome and plasmid.

The promoter diversity was characterized using the upstream sequences of 21,367 chromosome-associated and 1,242 plasmid-associated *bla*$_{TEM}$ genes. Multiple known promoters including P3, Pa/Pb, P4, and P5, which differ in transcriptional strength and expression level, were identified (19). Several novel promoter variants were detected, observing variations in the -10 and -35 elements. The promoter type distribution revealed distinct patterns between chromosomal and plasmid (Fig. 1d). It is important to note that the total number of plasmid sequences carrying *bla*$_{TEM}$ genes in this dataset was lower than the number of chromosome-associated *bla*$_{TEM}$ genes, which explains the reduced counts observed for plasmid promoter types. Despite this difference in sample size, clear patterns emerge. The P3 promoter is the most dominant type in both contexts, followed by Pa/Pb, while P4 and P5 were less frequent and were more commonly observed in chromosomal sequences. Notably, several novel -10/-35 promoter variants were detected predominantly among chromosome-associated *bla*$_{TEM}$ genes with only limited representation in plasmids. Overall, the chromosomal dataset exhibited broader promoter variation than the plasmid dataset; however, this difference may partly reflect the substantially larger number of chromosome-associated sequences analyzed.

Promoter-*bla*$_{TEM}$ variant relationships and phenotype-associated regulatory diversity

A decoupling analysis was performed to examine the relationship between promoter types and *bla*$_{TEM}$ variants in chromosomal and plasmid sequences. Additional file 2: Fig. S1 presents the promoter type distribution of individual *bla*$_{TEM}$ variants in chromosome (blue circles) and plasmid (orange triangles). Each panel represents one *bla*$_{TEM}$ variant, with the x-axis showing promoter categories (numeric codes) and the y-axis showing the number of sequences. The data show that while most minor or less-frequent *bla*$_{TEM}$ variants remain fixed to a single promoter type, whereas major high-frequency variants exhibit regulatory multi-mapping. P3

was the most frequently observed promoter across many *bla*$_{TEM}$ variants, especially in chromosomal sequences such as *bla*$_{TEM-1}$, *bla*$_{TEM-135}$, *bla*$_{TEM-12}$, and *bla*$_{TEM-122}$, confirming its central role in maintaining stable and efficient gene expression. Notably, ancestral variants like *bla*$_{TEM-1}$ display striking multi-promoter usage, shifting across multiple distinct promoter types within identical coding sequences. This suggests that the occurrence of different promoter types was independent of coding sequence evolution and may influence expression levels without altering enzyme structure. Importantly, some *bla*$_{TEM}$ variants, mainly in chromosomes, were observed with novel -10/-35 combinations.

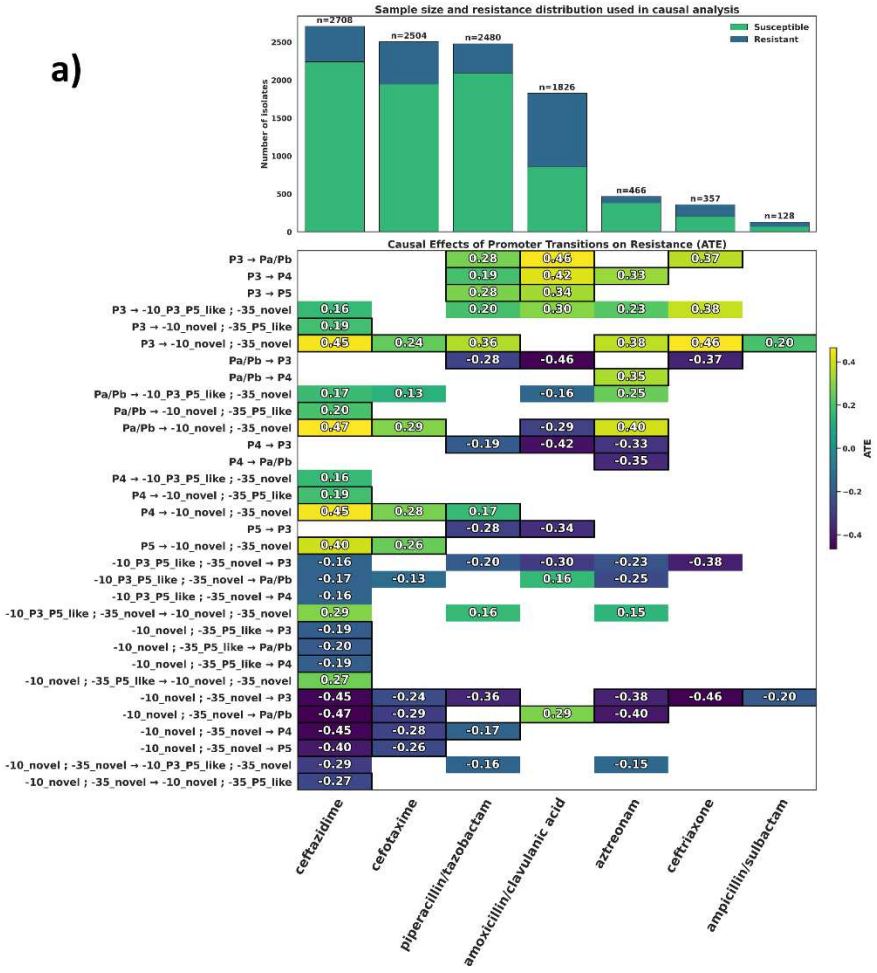
The analysis was further extended by integrating the available resistance phenotypes across multiple β-lactam antibiotics (Additional file 2: Fig. S2). P3 remained the predominant promoter and was most frequently associated with *bla*$_{TEM-1}$ across the antibiotics examined, highlighting its central role as the primary functional promoter driving *bla*$_{TEM}$ expression. Resistant isolates exhibited greater regulatory diversity compared to susceptible isolates, including the presence of stronger promoters such as Pa/Pb, P4 and P5, which are known to enhance transcriptional activity. Importantly, several novel promoter variants were identified, including -10_novel;-35_novel, -10_novel;-35_P5_like, -10_P3_P5_like;-35_novel, -10_novel;-35_P3_like, -10_P5_like;-35_novel. These novel promoter configurations were observed predominantly in resistant isolates and were rare or absent in susceptible populations.

Distribution and positional association of insertion sequences around *bla*$_{TEM}$ loci

To characterize the association of *bla*$_{TEM}$ loci with insertion sequences (IS), ISEScan was applied to the predefined 1-kb upstream and downstream regions surrounding *bla*$_{TEM}$ genes in chromosome- and plasmid-associated sequences. IS elements were identified in 2,948 of 20,317 *bla*$_{TEM}$ -positive chromosome isolates and 252 of 965 *bla*$_{TEM}$ -positive plasmid isolates. Based on ISEScan annotation, 30% of chromosome-associated IS elements were classified as complete and 70% as partial, whereas plasmid-associated IS elements comprised 73% complete and 27% partial elements. Additional file 2: Fig. S3 shows the distribution of the identified IS families according to their positional categories relative to the *bla*$_{TEM}$ locus. Overall, 17 different IS families were identified across 26 *bla*$_{TEM}$ variants in chromosome-associated sequences, while 13 IS families were identified across 11 *bla*$_{TEM}$ variants in plasmid-associated sequences. Among the 3,130 IS-associated *bla*$_{TEM}$ entries in chromosomes, P3 was the predominant promoter type (n = 2,951), followed by Pa/Pb (n = 91), −10_P3_P5_like; −35_novel (n = 54), P4 (n = 20), −10_novel; −35_novel (n = 8), −10_novel; −35_P5_like (n = 3), and P5 (n = 3). IS91 (n = 1,775), IS1182 (n = 614), and IS6 (n = 447) were the three most frequently identified IS families in the chromosome-associated dataset. Among the 324 IS-associated *bla*$_{TEM}$ entries in plasmids, P3 was similarly predominant (n = 306), followed by Pa/Pb (n = 6), P4 (n = 5), −10_novel; −35_novel (n = 5), and −10_P3_P5_like; −35_novel (n = 2). The most frequently identified plasmid-associated IS families were IS6 (n = 161), IS91 (n = 51), and IS1182 (n = 51). Additional file 1: Table S2 summarizes IS elements overlapping the upstream region of the *bla*$_{TEM}$ locus. Most upstream IS overlaps were located more than 400 bp from the gene start site, indicating that direct overlap with the proximal −35 and −10 promoter regions were comparatively limited. Additional file 2: Fig. S4a, S4b further shows the strand-normalized positional distribution of IS families relative to the *bla*$_{TEM}$ gene across the different promoter types in chromosome- and plasmid-associated sequences.

Counterfactual analysis of antimicrobial resistance difference across novel promoters

Among the 21,367 *bla*$_{TEM}$-associated promoter sequences analyzed, 494 (2.3%) did not match any of the established canonical promoter types P3, Pa/Pb, P4, or P5 (Additional file 1: Table S3). Interestingly, the presence of these non-canonical promoter types highlights under characterized regulatory diversity and may reflect evolutionary diversification within the *bla*$_{TEM}$ upstream region. Given the established differences in activity among canonical *bla*$_{TEM}$ promoters, we examined whether the novel promoter types were associated with differences in the predicted probability of antimicrobial resistance. At the chromosome level, we found 3327 genome isolates with resistance phenotypes and their promoter type information. None of the plasmid isolates were identified with resistance phenotype; therefore, the causal analysis was performed only on the chromosome entries. A counterfactual causal-inference framework was used to estimate promoter-associated differences between canonical and novel promoter types while maintaining for the constant *bla*$_{TEM}$ variant profiles. At the promoter-type level, transitions from the canonical P3 promoter to other promoter types showed positive ATE values across multiple antibiotics among comparisons meeting the predefined stability and permutation criteria. This pattern was observed for transitions from P3 to both canonical and several novel promoter types. Among the retained comparisons, the estimated ATEs for the P3-to-Pa/Pb, P3-to-P4, and P3-to-P5 transitions ranged from 0.24 to 0.46, 0.19 to 0.42, and 0.17 to 0.38, respectively. Transitions from P3 to novel promoter types also showed positive ATE values across multiple antibiotics, with estimates ranging from 0.16 to 0.46. The maximum estimated ATE was 0.46, representing a 46-percentage-point difference in model predicted resistance probability between the two counterfactual promoter conditions. For promoter pairs in which both transition directions were reported, the reverse contrasts showed corresponding ATE magnitudes with opposite signs, as expected from the ordered definition of the pairwise comparison (Fig. 2a). To assess whether the positive ATE pattern observed at the promoter-type level was shared across all novel motifs or was driven by only a subset with larger estimated effects, the analysis was subsequently extended to the individual promoter-motif level by performing counterfactual transitions between them. The analysis focused on transitions from canonical promoter motifs to novel motifs. Several canonical-to-novel motif transitions showed positive ATE values across multiple antibiotics, broadly resembling the pattern observed at the promoter-category level (Additional file 2: Fig. S5). Transitions originating from the canonical P3 motif showed a recurrent pattern of positive ATE values with multiple novel motifs across different antibiotics. Therefore, a focused analysis was performed to examine P3-to-other motif transitions (Fig. 2b). Several P3-to-novel motif transitions showed positive model-estimated ATEs across multiple antibiotics, although their magnitudes varied among the individual novel motifs.


a)
Sample size and resistance distribution used in causal analysis
Susceptible
Resistant
Number of isolates
n=2708
n=2504
n=2480
n=1826
n=466
n=357
n=128
Causal Effects of Promoter Transitions on Resistance (ATE)
P3 → Pa/Pb
P3 → P4
P3 → P5
P3 → -10_P3_P5_like ; -35_novel
P3 → -10_novel ; -35_P5_like
P3 → -10_novel ; -35_novel
Pa/Pb → P3
Pa/Pb → P4
Pa/Pb → -10_P3_P5_like ; -35_novel
Pa/Pb → -10_novel ; -35_P5_like
Pa/Pb → -10_novel ; -35_novel
P4 → P3
P4 → Pa/Pb
P4 → -10_P3_P5_like ; -35_novel
P4 → -10_novel ; -35_P5_like
P4 → -10_novel ; -35_novel
P5 → P3
P5 → -10_novel ; -35_novel
-10_P3_P5_like ; -35_novel → P3
-10_P3_P5_like ; -35_novel → Pa/Pb
-10_P3_P5_like ; -35_novel → P4
-10_P3_P5_like ; -35_novel → -10_novel ; -35_novel
-10_novel ; -35_P5_like → P3
-10_novel ; -35_P5_like → Pa/Pb
-10_novel ; -35_P5_like → P4
-10_novel ; -35_P5_like → -10_novel ; -35_novel
-10_novel ; -35_novel → P3
-10_novel ; -35_novel → Pa/Pb
-10_novel ; -35_novel → P4
-10_novel ; -35_novel → P5
-10_novel ; -35_novel → -10_P3_P5_like ; -35_novel
-10_novel ; -35_novel → -10_novel ; -35_P5_like
ATE
ceftazidime
cefotaxime
piperacillin/tazobactam
amoxicillin/clavulanic acid
aztreonam
ceftriaxone
ampicillin/sulbactam

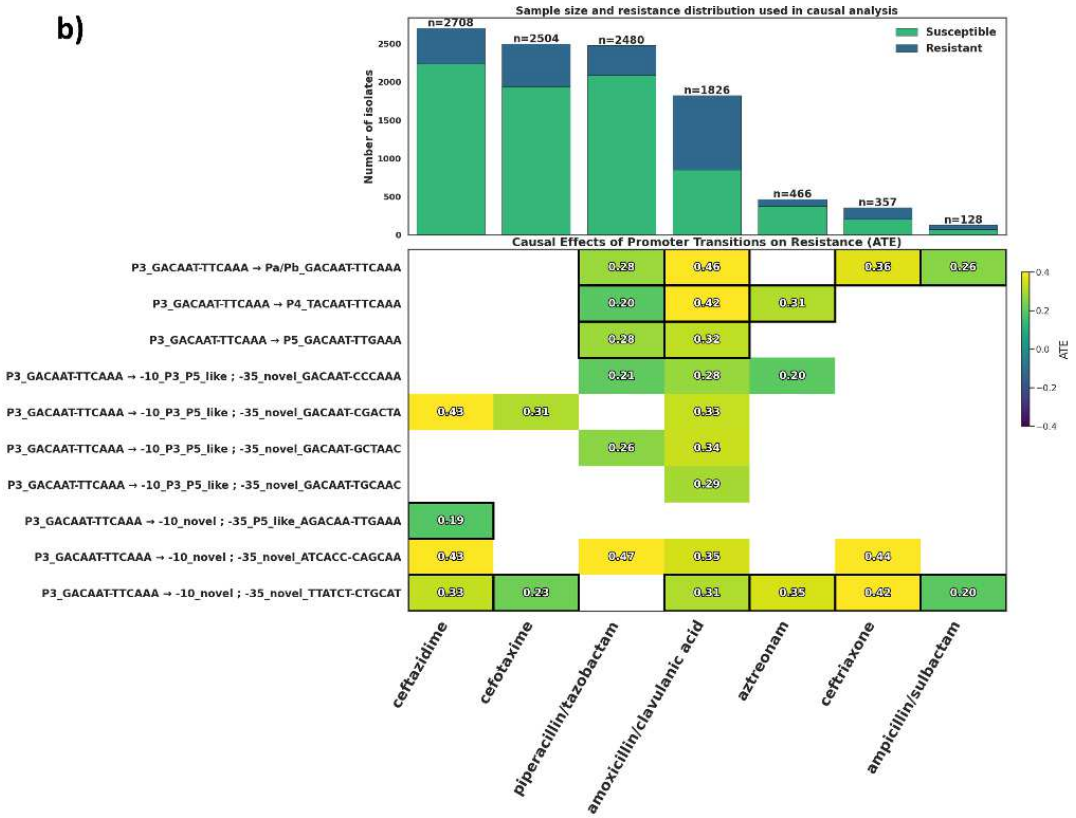

b)
Sample size and resistance distribution used in causal analysis
Susceptible
Resistant
Number of isolates
n=2708
n=2504
n=2480
n=1826
n=466
n=357
n=128
Causal Effects of Promoter Transitions on Resistance (ATE)
P3_GACAAT-TTCAAA → Pa/Pb_GACAAT-TTCAAA
P3_GACAAT-TTCAAA → P4_TACAAT-TTCAAA
P3_GACAAT-TTCAAA → P5_GACAAT-TTGAAA
P3_GACAAT-TTCAAA → -10_P3_P5_like ; -35_novel_GACAAT-CCCAAA
P3_GACAAT-TTCAAA → -10_P3_P5_like ; -35_novel_GACAAT-CGACTA
P3_GACAAT-TTCAAA → -10_P3_P5_like ; -35_novel_GACAAT-GCTAAC
P3_GACAAT-TTCAAA → -10_P3_P5_like ; -35_novel_GACAAT-TGCAAC
P3_GACAAT-TTCAAA → -10_novel ; -35_P5_like_AGACAA-TTGAAA
P3_GACAAT-TTCAAA → -10_novel ; -35_novel_ATCACC-CAGCAA
P3_GACAAT-TTCAAA → -10_novel ; -35_novel_TTATCT-CTGCAT
ATE
ceftazidime
cefotaxime
piperacillin/tazobactam
amoxicillin/clavulanic acid
aztreonam
ceftriaxone
ampicillin/sulbactam

Fig. 2 Counterfactual effects of *bla*$_{TEM}$ promoter transitions on antimicrobial resistance estimated using average treatment effects (ATEs). (a) Promoter-type-level transitions among canonical and non-canonical *bla*$_{TEM}$ promoter types across the β-lactam antibiotics. (b) Promoter-motif-level transitions originating from the canonical P3 motif to other canonical and novel motifs, highlighting the recurrent positive ATE pattern observed for P3-originating transitions. In both panels, ATEs were estimated while holding the observed *bla*$_{TEM}$ variant profile constant, with each row representing the ordered transition indicated by the arrow. Positive ATEs, shown from green to yellow with increasing magnitude, indicate a positive difference in model-predicted resistance probability under the second promoter condition relative to the first, whereas negative ATEs, shown from blue to purple, indicate a negative difference. Only promoter-antibiotic transitions meeting the predefined criteria of ATE stability ≥ 0.80 and permutation ($p \leq 0.05$) are shown; all others are left blank. Bold black borders indicate transitions with a non-zero pairwise overlap fraction; displayed cells without borders lacked empirical overlap between the compared promoter conditions and therefore represent model-based counterfactual estimates. The upper stacked bars show the number of isolates with available phenotype data for each antibiotic and their susceptible (green) and resistant (blue) distributions.

Resistance phenotype and genomic characterization of putative novel promoter motifs in single-*bla*$_{TEM}$ isolates

We identified several non-canonical promoter motifs with sequence variations in the -35 and -10 regions upstream of chromosomally located *bla*$_{TEM}$ genes. To assess their potential functional and phenotypic relevance, we specifically analyzed isolates carrying a single *bla*$_{TEM}$ variant with novel promoter motifs and with the absence of any additional β-lactamase genes. The identified novel promoters were mapped to their corresponding isolate metadata and antibiotic susceptibility profiles. A total of 12 such isolates were obtained, each exhibiting phenotypic resistance to at least one β-lactam antibiotic. Promoter sequence analysis of these isolates revealed five distinct novel promoter motifs. Table 1 summarizes their frequencies among the resistant isolates and across the entire *bla*$_{TEM}$ containing isolates. Among the 12 isolates for which metadata were available, six distinct multilocus sequence types (MLSTs) were identified, with MLST.ecoli_achtman_4.131 represented in five isolates. The isolates were collected over a broad temporal range (2003–2017) and originated from four countries such as the United Kingdom, Norway, Brazil, and the United States (Additional file 1: Table S4). The selected isolates carried either *bla*$_{TEM\text{-}1}$ or *bla*$_{TEM\text{-}135}$, with each isolate containing a single novel promoter motif. Analysis of the available MIC and disk-diffusion data showed that isolates carrying the four *bla*$_{TEM\text{-}1}$ associated motifs and the single *bla*$_{TEM\text{-}135}$ associated motif were resistant to ampicillin and piperacillin/tazobactam. Among the *bla*$_{TEM\text{-}1}$ isolates, three motifs- M1, M2, and M5 contained the novel -35 elements CCCAAA, TGCAAC, and TTGAAG, respectively, together with the canonical P3/P5-like -10 motif GACAAT. The remaining motifs were M4 in a *bla*$_{TEM\text{-}1}$ isolate (-10: TGCAAC; -35: TAAGTT) and M3 in a *bla*$_{TEM\text{-}135}$ isolate (-10: AAGTAA; −35: TAAAAG). Across the complete dataset, M1 was detected in 97 entries, followed by M2 in 96 entries and M4 in six entries, whereas M3 and M5 were each detected in a single entry. Based on their recurrence in the complete dataset, three motifs- M1, M2, and M4 were selected for structural analysis to examine their potential interactions with the σ^{70} subunit.

Table 1: Putative novel promoter motifs identified among resistant single-*bla*$_{TEM}$ isolates and their occurrence in the complete dataset. The table summarizes the corresponding -10 and -35 motif sequences, the number of single-*bla*$_{TEM}$ isolates having resistance to at least one β-lactam antibiotic, and the frequency of occurrence of each motif across the complete dataset.

Motif Number	Novel Promoter Type	-10 Region	-35 Region	Resistant isolates count with single-*bla*$_{TEM}$ gene	Frequency of occurrence in data
Motif 1	-10_P3_P5_like; -35_novel	GACAAT	CCCAAA	3	97
Motif 2	-10_P3_P5_like; -35_novel	GACAAT	TGCAAC	5	96
Motif 3	-10_novel; -35_novel	AAGTAA	TAAAAG	1	1
Motif 4	-10_novel; -35_novel	TGCAAC	TAAGTT	2	6
Motif 5	-10_P3_P5_like; -35_novel	GACAAT	TTGAAG	1	1

Interaction patterns of σ^{70} with canonical and putative novel promoter motifs

To compare the protein-DNA interaction profiles of canonical and novel promoter motifs, the modelled σ^{70}-DNA complexes were subjected to 100 ns molecular dynamics simulations. The trajectories were analyzed for hydrogen-bond, salt-bridge, π-cation, and π–π stacking interactions, and contacts present in at least 10% of the trajectory frames were retained for comparison. Six-base motifs corresponding to the -35 and -10 regions were examined at their respective σ^{70} interaction interfaces. The -35 region was analyzed in a double-stranded DNA context, whereas the -10 region was analyzed in a single-stranded DNA context. At the σ^{70} domain 4-DNA interface, the canonical -35:P3/P4-like (TTCAAA) and -35:P5-like (TTGAAA) motifs exhibited a total of 21 and 32 interactions, respectively. The novel -35:M1 (CCCAAA), -35:M2 (TGCAAC), and -35:M4 (TAAGTT) motifs also exhibited 16, 25, and 25 interactions, respectively. Thus, M2 and M4 retained more persistent contacts than the canonical P3/P4-like motif, whereas M1 showed the lowest number of interactions. The P3/P4-like motif contained 12 hydrogen bonds and nine salt bridges, while the P5-like motif contained 13 hydrogen bonds and 19 salt bridges. Among the novel motifs, -35:M1, -35:M2, and -35:M4 exhibited 10 and 6, 13 and 12, and 11 and 14 hydrogen-bond and salt-bridge interactions, respectively (Table 2). These findings indicate that -35 sequence variation was associated with differences in both the number and composition of protein–DNA contacts. In particular, the greater number of contacts in the P5-like motif was mainly associated with an increased contribution of salt bridges, while -35:M4 also showed more salt bridges despite having fewer hydrogen bonds than the P3/P4-like motif. Residue-level analysis identified 12 σ^{70} residues (Arg554, Arg562, Leu573, Arg584, Arg586, Arg588, Gln589, Lys593, Arg596, Lys597, Arg599, His600) participating in hydrogen-bond and/or salt-bridge interactions across the five -35 simulations (Additional file 1: Table S5a). Although several residues were shared among motifs, the greatest overlap was observed between the canonical -35:P5-like motif and

novel -35:M2 motif, which shared seven interacting residues. Overall, the motifs engaged different subsets of σ^{70} residues while retaining a smaller group of recurrent contacts, indicating promoter-dependent variation in the residue-level interaction network. At the σ^{70} domains 2-DNA interface, the canonical -10:P3/P5-like (GACAAT) and -10:P4-like (TACAAT) motifs exhibited a total of 20 and 17 interactions, respectively, while the novel -10:M4 (TGCAAC) motif exhibited 18 interactions. Hydrogen bonds were the predominant interaction class, with 14, 13, and 14 interactions observed for the -10:P3/P5-like, -10:P4-like, and -10:M4 novel motifs, respectively. Each canonical motif formed two salt bridges, whereas the novel motif formed three. One π-cation interaction was detected for each motif. In contrast, π–π stacking interactions were observed only for the canonical motifs, with three interactions for -10:P3/P5-like and one for -10:P4-like, while none persisted for the novel motif. Across the three -10 promoter simulations, 17 distinct protein residues (Ala382, Asn383, Arg385, Leu386, Lys418, Glu420, Arg422, Arg423, Tyr425, Lys426, Ser428, Thr429, Tyr430, Thr432, Trp433, Trp434 and Gln437) were involved with varying interaction occupancies and the residue composition among motifs (Additional file 1: Table S5b). Overall, the novel -10 and -35 motifs exhibited interactions with σ^{70}, with localized differences in interaction number, type, residue participation, and occupancy relative to the canonical motifs. Based on the interactions and frequency analysis observed in the putative novel motif -35:M2 showed a consistent pattern over the simulation trajectory and their 3d-interaction view was reported in Additional file 2: Fig. S6.

Table 2: Motif-level σ^{70}-DNA interaction counts retained at ≥10% occupancy during the 100 ns molecular dynamics simulations. Interaction counts are shown separately for the -35 and -10 promoter regions and include hydrogen bonds, salt bridges, π–π interactions, and π–cation interactions. The total represents the sum of all interaction types meeting the predefined ≥10% trajectory occupancy threshold for each canonical or novel motif.

Region	Motif	H-bond	Salt bridge	π–π	π–cation	Total
-35 region	P3/P4-like (TTCAAA)	12	9	0	0	21
	P5-like (TTGAAA)	13	19	0	0	32
	CCCAAA-M1	10	6	0	0	16
	TGCAAC-M2	13	12	0	0	25
	TAAGTT-M4	11	14	0	0	25
-10 region	P3/P5-like (GACAAT)	14	2	3	1	20
	P4-like (TACAAT)	13	2	1	1	17
	TGCAAC-M4	14	3	0	1	18

Discussion

TEM β-lactamases encoded by *bla*$_{TEM}$ and its variants collectively hydrolyze a wide range of β-lactam antibiotics, thereby contributing to antimicrobial resistance. Although bacteria have developed multiple resistance mechanisms, the level of resistance-gene expression plays an important role in determining the resistance phenotype and is modulated by upstream regulatory regions. This study focuses on understanding the regulatory evolution of *bla*$_{TEM}$

genes and its effect on the antimicrobial resistance in *E. coli* using large scale global genome analyses.

The *E. coli* isolates analyzed in this study were MDR and harbored multiple AMR genomic factors. This study focuses on understanding the evolution of regulatory regions of the *bla*$_{TEM}$ genes along with its variants. The chromosome-derived dataset contained a greater number of distinct *bla*$_{TEM}$ variants than the plasmid-derived dataset. The proportion of *bla*$_{TEM}$ -positive sequences was approximately 38% in the plasmid-derived dataset and 34% in the chromosome-derived dataset. This distribution is broadly consistent with previous studies highlighting the extensive occurrence and exchange of antimicrobial resistance genes across plasmid and chromosomal compartments (21,22). These findings are consistent with the important role of plasmids as mobile genetic elements facilitating the dissemination of antimicrobial resistance genes, whereas chromosomally encoded genes are primarily inherited vertically but can also undergo horizontal transfer through mobile genetic elements (17,23–25).

The β-lactam resistance phenotypes of *bla*$_{TEM}$-variant-carrying isolates were examined to characterize their susceptibility patterns. Phenotypic information was available for 17 β-lactam antibiotics. The occurrence of *bla*$_{TEM}$ in both resistant and susceptible isolates indicate that gene presence alone does not consistently correspond to the resistance phenotype, suggesting that additional factors, including the regulatory context of the gene, may contribute to the observed phenotype. The detection of *bla*$_{TEM}$ genes in both resistant and susceptible isolates align with large-scale genomic studies showing that antimicrobial resistance is shaped by complex interactions among resistance genes, mutations, and regulatory elements, rather than gene presence alone. Experimental studies have shown that reduced or heterogeneous expression of resistance determinants can result in differences in β-lactam susceptibility despite the presence of the corresponding resistance gene (26–28). Complementing these experimental observations, recent ML/DL-based genomic studies have shown that resistance phenotypes are better explained by combinations of resistance genes, mutations, and other genomic features than by gene carriage alone, further highlighting the potential importance of regulatory context in interpreting AMR phenotypes (29–31). Together, these results highlight the potential importance of regulatory mechanisms, including promoter regions, in modulating the expression of resistance determinants and contributing to differences in antimicrobial susceptibility. Thus, analysis of upstream regulatory regions alongside resistance-gene presence may provide a more complete understanding of the mechanisms underlying antimicrobial resistance.

Given the importance of regulatory-region analysis in the context of AMR genes, this study first examined the evolutionary conservation and regulatory-region diversity of *bla*$_{TEM}$ genes in chromosomal and plasmid contexts. On observing distal upstream positions, variability was more prominent, suggesting greater evolutionary modification of the upstream sequence, which contains regulatory elements such as promoters and transcription-factor binding sites. Both chromosome- and plasmid-associated *bla*$_{TEM}$ sequences showed a similar overall Shannon entropy pattern, with greater variability in the upstream region than in the coding region. When comparing the upstream regions between the two genomic contexts, plasmid-associated sequences showed greater variability than chromosome-associated sequences. This observation is consistent with the dynamic nature of plasmids as mobile genetic elements, which frequently undergo horizontal transfer, recombination, and mobile-element-mediated rearrangements that may contribute to diversification of their surrounding genetic regions (32,33). In contrast, the

comparatively lower variability observed in chromosome-associated upstream regions may reflect greater sequence constraint within the chromosomal context, consistent with the relatively stable organization and evolutionary constraints operating within bacterial chromosomes (34,35). Existing studies have shown that variation within the upstream regulatory and genetic environment of *bla*$_{TEM}$ can influence gene expression and contribute to differences in β-lactam resistance phenotypes. In *bla*$_{TEM}$, established promoter types differ in their -35 and -10 motifs and have been associated with differences in *bla*$_{TEM}$ expression, β-lactamase activity, and β-lactam resistance (19,36,37). The relatively low entropy observed around these core promoter positions is therefore consistent with functional constraint on promoter-associated sequence elements, although sequence variation is still present within these regions. The comparatively low entropy observed around these core promoter positions is therefore consistent with functional constraint on promoter-associated sequence elements, although sequence variation is still present within these regions. The coding region shows very low entropy in both chromosome and plasmid, indicating that most of the *bla*$_{TEM}$ coding sequence remains highly conserved despite the emergence of numerous *bla*$_{TEM}$ variants, Nevertheless, functionally important amino-acid substitutions can occur at particular positions and alter substrate recognition and catalytic properties, contributing to the diversification of *bla*$_{TEM}$ β-lactamases(10,12). Overall, this entropy plot reinforces that the *bla*$_{TEM}$ coding region is highly conserved and functionally constrained, while the upstream regulatory region exhibits greater variability, reflecting evolutionary flexibility in gene expression regulation alongside comparatively limited changes in the coding sequence. A focused analysis of the upstream regulatory region was performed by identifying the -35 and -10 promoter elements relative to the *bla*$_{TEM}$ coding start site, which are involved in σ-factor recognition during transcription initiation. Based on these motifs, the identified regions were categorized into established canonical promoter types such as P3, Pa/Pb, P4, and P5. These promoters have previously been characterized and are associated with different levels of *bla*$_{TEM}$ expression and β-lactamase activity (19). The analysis of diverse promoter types reveals P3 as the predominant promoter type in both chromosomes- and plasmid-associated sequences, consistent with the historical positioning as the ancestral *bla*$_{TEM}$ promoter. Early genetic mapping and transposon analyses have established P3 as the native regulatory element associated with the classical Tn3 and Tn2 transposons (16,38,39). The analyses also identified several novel promoter combinations in chromosomes and comparatively less in plasmid. From an evolutionary standpoint, one possible explanation for this difference may involve the contrasting gene-dosage contexts of plasmid- and chromosome-associated *bla*$_{TEM}$. Experimental studies using multicopy *bla*$_{TEM-1}$ carrying plasmids have shown that increased plasmid copy number can increase gene dosage, expression, and resistance, thereby providing an additional mechanism for achieving elevated resistance levels independent of promoter-sequence changes (23,40). In contrast, chromosome-associated *bla*$_{TEM}$ generally occurs in a lower-copy genomic context, where changes in promoter strength can provide an alternative mechanism for increasing gene expression and β-lactamase production under antibiotic selection. Promoter mutations in chromosomally encoded β-lactamases have experimentally been shown to substantially increase expression and resistance (19,20,41). This regulatory route may therefore help explain the greater diversity of non-canonical promoter motifs observed among chromosome-associated *bla*$_{TEM}$ sequences in the present study.

The overall distribution of identified promoter types was examined across *bla*$_{TEM}$ variants and resistance phenotypes, as shown in Additional file 2: Fig. S1 and S2. Within our dataset, P3

was the most abundant promoter type and was most frequently associated with *bla*$_{\text{TEM-1}}$, which occurred in both resistant and susceptible isolates. *bla*$_{\text{TEM-1}}$ was also the most frequently observed variant and was linked to multiple promoter types in both chromosome- and plasmid-associated sequences. These patterns are consistent with previous reports describing P3 as the ancestral promoter upstream of *bla*$_{\text{TEM-1}}$ and documenting the occurrence of *bla*$_{\text{TEM-1}}$ in different promoter backgrounds (19,42). This suggests that promoter diversification can alter *bla*$_{\text{TEM}}$ expression without changes to its coding sequence. This observation is consistent with the entropy analysis as discussed in the Fig. 1c, which demonstrated strong conservation across the *bla*$_{\text{TEM}}$ coding region while revealing substantially greater variability within upstream regulatory sequences. Together, these findings suggest that regulatory diversification can occur within *bla*$_{\text{TEM}}$ populations without requiring extensive changes to the β-lactamase coding sequence, with the newly identified promoter types potentially reflecting evolutionary diversification of the regulatory region. Comparison of promoter distributions between resistant and susceptible isolates revealed phenotype-associated differences across several β-lactam antibiotics. Although P3 remained the dominant promoter in both groups, other canonical promoter categories, such as Pa/Pb, P4, and P5, and with several novel -10 and -35 motif combinations, showed greater relative representation among resistant isolates for multiple antibiotics, although this pattern was not uniform across all antibiotics. This pattern suggests a possible contribution of promoter variation to differences in resistance phenotype and is consistent with the established influence of *bla*$_{\text{TEM}}$ promoter variation on *bla*$_{\text{TEM}}$ expression and β-lactam resistance (19,20,43). This regulatory and coding region decoupling highlights an alternative evolutionary route in which a narrow-spectrum *bla*$_{\text{TEM}}$ enzyme can contribute to increased β-lactam resistance through altered gene expression while retaining the same encoded *bla*$_{\text{TEM}}$ protein. Experimental and clinical studies have similarly shown that promoter variation and *bla*$_{\text{TEM-1}}$ hyperproduction can substantially alter β-lactam resistance without changing the enzyme itself (44,45). Similarly, the evolution of extended-spectrum β-lactamases and inhibitor-resistant *bla*$_{\text{TEM}}$ variants involves amino-acid substitutions that modify the functional properties of the *bla*$_{\text{TEM}}$ enzyme (10,19,46,47). The observation on the surrounding IS context reveals that a notable higher proportion of complete IS elements was observed in plasmids (73%) than in chromosomes (30%), consistent with the established role of plasmids as major carriers of mobile genetic elements involved in antimicrobial-resistance dissemination (17). However, the overall IS-family distribution was broadly similar between canonical and novel promoter contexts, suggesting no distinct IS signature associated with the novel motifs. IS6, IS91 and IS1182 were among the most frequently observed families where IS26 of the IS6 family has been directly linked to amplification and mobilization of *bla*$_{\text{TEM-1B}}$ in *E. coli*, while IS91- and ISCfr1/IS1182-associated genetic contexts containing *bla*$_{\text{TEM}}$ have also been reported (48,49). While most upstream IS overlaps were located beyond 400 bp from the gene start, overlaps were also detected within the proximal upstream regions of *bla*$_{\text{TEM}}$. The occurrence of IS elements in these regions is consistent with previous reports of IS-mediated modification of TEM promoter architecture, supporting the relevance of examining the surrounding IS context in promoter evolution that may modify the promoter architecture and influence the expression (50).

In this study, we examined whether counterfactual transitions from canonical to newly identified promoter types were associated with higher estimated probabilities of antimicrobial resistance after adjustment for the observed *bla*$_{\text{TEM}}$ variant profiles. This framework allowed the resistance probability under two alternative promoter conditions to be compared while

retaining the same *bla*$_{TEM}$ variant profile for each isolate. Causal and counterfactual approaches have also been applied or proposed in antimicrobial resistance research to examine outcomes under alternative antibiotic exposures, define appropriate counterfactuals for estimating the burden attributable to resistance, and evaluate antimicrobial treatment strategies (51–53). These studies highlight the value of causal reasoning in AMR for moving beyond observational associations to examine potential changes in outcomes under alternative exposures or interventions. At the promoter-category level, transitions from the canonical P3 promoter to other canonical and novel promoter types showed positive ATE values across multiple antibiotics. For the canonical promoter comparisons, the positive direction of the P3 to other canonical promoter estimates was broadly consistent with the established functional differences among *bla*$_{TEM}$ promoters reported by (19). This positive pattern was further examined at the motif level to determine whether it was shared across all novel motifs or limited to a subset. Several P3-to-novel motif transitions showed positive estimated effects across the analyzed antibiotics, suggesting that the differences may also extend to specific promoter motif level. Although these estimates should be interpreted cautiously, particularly for rare promoter architectures with limited promoter-*bla*$_{TEM}$ overlap, their consistency across multiple antibiotics and stability under bootstrap and permutation analyses support the use of this counterfactual framework for exploring potential relationships between promoter variation and resistance-associated phenotypic differences.

The causal analysis identified several novel promoter motifs with significant resistance-associated effects relative to the baseline P3 promoter. To reduce the influence of additional β-lactamase determinants, these motifs were subsequently examined in isolates carrying a single *bla*$_{TEM}$ variant and no other β-lactamase genes. Five novel promoter variants were identified among 12 phenotypically resistant isolates, of which M1(-10(GACAAT); -35(CCCAAA)), M2(-10(GACAAT); -35(TGCAAC)), and M4(-10(TGCAAC); -35(TAAGTT)), which recurred across the complete dataset were selected for structural analysis (Table 1 and Additional file 1: Table S4). To provide structural support for considering these sequences as putative promoter motifs, protein-DNA complexes were subjected to 100 ns molecular dynamics simulations using the corresponding σ^{70} promoter-recognition domains. The canonical -10 and -35 motifs associated with P3, P4, and P5 were first examined as reference systems and compared with experimentally established promoter characteristics and reported σ^{70}-promoter interactions, thereby providing a benchmark against which the protein-DNA interaction patterns of the novel motifs could subsequently be evaluated. Among the canonical -35 motifs, the P5-like sequence (TTGAAA) showed more retained σ^{70}-DNA interactions than the P3/P4-like sequence (TTCAAA; 32 versus 21), largely due to a greater contribution of salt bridges. This pattern is directionally consistent with Lartigue et al., who reported the highest *bla*$_{TEM-1}$ expression from the P5 promoter and noted that its -35 sequence more closely resembles the σ^{70} consensus element TTGACA. Their isogenic analysis further showed that promoter sequence variation can alter β-lactamase activity and β-lactam susceptibility (19). Although interaction number alone cannot be interpreted as a direct measure of promoter strength, this agreement supports the biological relevance of the canonical -35 interaction pattern. At the -10 interface, the P3/P5-like and P4-like motifs showed broadly comparable interaction profiles, with 20 and 17 retained contacts, respectively, predominantly involving hydrogen bonds. Residue-level analysis of the canonical -35 motif simulations was further compared with previously reported σ^{70}-promoter interactions. Of the 11 σ^{70} residues involved in the canonical interaction profiles, seven, Arg554, Arg562, Leu573, Arg584, Arg586,

Arg588, and Gln589 have previous structural or experimental support for involvement in -35 recognition or promoter-DNA interaction (54–56). In particular, genetic and structural studies have implicated Arg584 and Arg588 in sequence-specific recognition of the -35 element, with Arg584 directly contacting the -31 base pair and Arg588 influencing promoter specificity at the -33 position (54,57). Evidence for Lys593 has been reported to contact the promoter-DNA backbone upstream of the -35 hexamer, around positions -38 to -39, although its specific contribution to sequence-specific -35 recognition is less clearly established (58), while Arg596, Lys597, and Arg599 are better characterized in other σ^{70} region-4 interaction contexts than as established -35 DNA-contact residues (55,56,59,60). Their DNA contacts observed in the present simulations were therefore treated as additional simulation-derived interactions. Residue-level analysis of the canonical -10 motif simulations was similarly compared with previously reported σ^{70} region-2-promoter interactions. 14 distinct residues were identified in the simulations. Several residues identified in the canonical -10 simulations, including Asn383, Arg385, Leu386, Lys418, Glu420, Ser428, Thr429, Tyr430, Thr432, and Trp433 correspond to conserved σ-factor positions previously implicated in −10 promoter recognition by structural studies (61) with several of these residues additionally supported by *E. coli*-specific structural or mutational studies (62,63). Among these, Thr429 plays an important role in promoter DNA melting, whereas Tyr430 participates in recognition of the flipped -11A base through aromatic stacking (64,65). Consistent with this mechanism, π–π stacking involving Tyr430 and promoter DNA was also observed in the canonical -10 simulations. Lys426 has been implicated in interactions with the −10 promoter region (62), while Trp433 and Trp434 form part of the aromatic region involved in single-stranded DNA binding and open-complex formation (63,66). Structural studies further support sequence-specific recognition of conserved −10 bases, including T−7, by σ region 2 (67). Gln437 has additionally been linked genetically to sequence recognition at the -12 position (68). For Ala382, clear residue-specific evidence for an established -10 DNA-contact role was not identified in the primary literature examined; its interaction was therefore considered an additional simulation-derived contact. Having established the canonical interaction profiles, the selected novel motifs were examined using the same criteria. Among the novel -35 motifs, M1 and M2 exhibited 16 and 25 retained interactions, respectively, involving a combined set of 10 σ^{70} residues, seven of which had previous structural or experimental support for promoter-DNA interaction. The -35 motif of M4 similarly exhibited 25 interactions involving seven residues, although only three had comparable literature support. At the -10 interface, the -10 motif of M4 formed 18 interactions involving 11 residues, 10 of which had previously reported structural or functional relevance to σ^{70}-promoter recognition. Thus, the novel motifs retained σ^{70} interaction networks while differing in the number, type, and residue composition of their contacts relative to the canonical motifs. The novel motifs were therefore evaluated using three complementary lines of evidence: their recurrence across the genomic dataset, the presence of protein-DNA interactions exceeding the predefined occupancy threshold, and the involvement of σ^{70} residues with previous structural or experimental support for promoter-DNA recognition. Considering these criteria together allowed recurrent sequence variants with substantial interaction networks involving multiple literature-supported residues to be distinguished from less-supported variants. Among the -35 motifs, M2 (TGCAAC) emerged as the best-supported motif based on the combined criteria, occurring in 96 entries and maintaining 25 interactions involving several residues previously associated with σ^{70}-DNA recognition. Together with its

estimated effect on the resistance phenotype relative to P3 in the causal analysis, these observations support the consideration of M2 as a putative -35 regulatory motif.

The findings of the present study should be considered in light of a few limitations. Although *bla*$_{TEM}$ variants were identified using a sequence similarity threshold of >90%, variant-specific differences in resistance strength were not independently categorized. Novel promoter motifs were identified based on their positional correspondence with previously reported promoter motifs and a fixed 17-bp spacer between the -35 and -10 regions. Although bootstrap resampling and permutation testing were used to assess the robustness of the causal estimates, these estimates remain dependent on the underlying data structure, where sufficient overlap between promoter transitions was limited, the resulting estimates may involve extrapolation, and therefore require cautious interpretation. In addition, although the single-*bla*$_{TEM}$-isolate analysis excluded isolates carrying additional β-lactamase genes, other resistance mechanisms, including efflux systems and porin alterations, may also contribute to the observed phenotype. Despite these limitations, the integration of genomic occurrence patterns, causal analysis, and molecular dynamics-based protein-DNA interaction profiles provide complementary evidence for the contribution of regulatory-region variation to *bla*$_{TEM}$-associated resistance. The agreement between the interaction profiles of selected novel promoter motifs and established structural and experimental evidence for σ^{70}-promoter recognition further supports their potential functional relevance and provides a basis for their consideration as putatively evolved promoter motifs. Further experimental characterization would enable direct assessment of their promoter activity and contribution to resistance.

Conclusion

This large-scale genomic analysis of the regulatory regions of *E. coli bla*$_{TEM}$ genes highlights the importance of investigating promoter variation as a potential contributor to differences in antimicrobial resistance phenotypes. Exploratory analysis shows the contrasting patterns of sequence conservation between the highly conserved *bla*$_{TEM}$ coding region and its variable upstream regulatory regions. The promoter diversity revealed broadly similar distributions of promoter types across chromosomal- and plasmid-associated *bla*$_{TEM}$ genes, while also identifying diverse novel promoter configurations. Causal analysis identified multiple novel promoter motifs associated with significantly greater resistance effects compared with the canonical P3 promoter. Subsequent evaluation in single-*bla*$_{TEM}$ containing isolates and molecular dynamics-based structural analysis demonstrated that the selected novel motifs maintained interactions with their respective σ^{70} domains. Among these, M2 (TGCAAC) at the -35 region emerged as the best-supported candidate for consideration as a putatively evolved promoter motif based on combined genomic, casual and structural evidence. Overall, this study demonstrates the utility of integrating large-scale genomic analysis, causal inference, and structural approaches to investigate regulatory-region evolution and provides insights into the potential contribution of promoter variation in altering gene expression and contributing to antimicrobial resistance.

Methods

Genome data collection and resistance gene identification

The BV-BRC database was systematically screened for *E. coli* genomes that were classified as whole-genome sequenced (WGS) or complete assemblies and annotated with genome quality

as Good. Based on these criteria, a total of 60,087 *E. coli* genome FASTA files were downloaded as of 10 December 2025 (69). AMR gene screening was performed using ABRicate v1.0.1 (70) against the Comprehensive Antibiotic Resistance Database (CARD) (71). The locally updated CARD database used for this analysis contained 6,052 curated AMR genes as of the database update performed on February 1, 2026. Thresholds of 90% sequence identity and 90% coverage were applied for gene detection.

Identification and extraction of upstream regulatory regions of *bla*$_{\text{TEM}}$ variants and IS elements

Using custom python scripts, genomic contigs were classified into chromosomal and plasmid sequences and exported into separate FASTA files. This process resulted in 60,087 chromosome FASTA files and 2,548 plasmid FASTA files. An in-house python script was then used to extract upstream regulatory regions for isolates carrying *bla*$_{\text{TEM}}$ β-lactamase variants. For each TEM-positive isolate, upstream regions were extracted with a minimum length of 100 bp and extended up to 500 bp where sequence context permitted. This analysis identified 20,317 genome isolates carrying a total of 21,144 *bla*$_{\text{TEM}}$ variants located on chromosomes, and 965 genome isolates carrying 1,242 *bla*$_{\text{TEM}}$ variants on plasmids (Fig. 1a). Promoter regions were subsequently screened using in-house python scripts based on promoter definitions described in the reference study (19). Isolates lacking known promoter motifs in the canonical -10 and -35 regions were analyzed separately to identify putative novel promoter patterns. The most frequently occurring -35 and -10 motif coordinates observed among the canonical promoters were used as positional references, while a default 17-bp spacer was used to define the relative spacing between the -35 and -10 elements. ISEscan (v1.7.3) was used to extract the IS elements on these *bla*$_{\text{TEM}}$ -containing isolates (72). An inhouse script was used to extract IS elements surrounding *bla*$_{\text{TEM}}$ genes within 1000 bp and classified as inside, near and overlap. The inside category was defined when the entire AMR gene was located within the start and end coordinates of an identified IS element. The near category corresponded to IS elements located within 1000 bp surrounding the *bla*$_{\text{TEM}}$ gene without overlapping its coordinates, whereas the “overlap” category corresponded to IS elements whose coordinates partially overlapped with the *bla*$_{\text{TEM}}$ gene coordinates.

Curation and preprocessing of antimicrobial resistance phenotype data

For all selected isolates, laboratory-confirmed AMR phenotypes, including resistant phenotypes were retrieved from the associated metadata. These were further curated to extract resistance and susceptibility profiles for the β-lactam antibiotics. However, based on the availability, we identified 17 β-lactam antibiotics which can be cleaved by *bla*$_{\text{TEM}}$ variants, including amoxicillin, amoxicillin-clavulanic acid, ampicillin, ampicillin-sulbactam, aztreonam, carbenicillin, cefazolin, cefotaxime, cefpodoxime, cefpodoxime-clavulanic acid, ceftazidime, ceftriaxone, cephalothin, penicillin, piperacillin, piperacillin-tazobactam, and ticarcillin-clavulanic acid. During data preprocessing, intermediate resistance phenotypes were conservatively reclassified as resistant and other β-lactam antibiotics which do not have proper phenotype data were excluded due to missing phenotype annotations.

Counterfactual analysis of promoter-associated effects on antimicrobial resistance

The contribution of promoter variation to antimicrobial resistance was investigated using a regression-based counterfactual causal inference workflow. Promoter type was considered the

treatment variable, antibiotic resistance phenotype the binary outcome, and the presence or absence of individual *bla*$_{\text{TEM}}$ variants was incorporated as observed adjustment covariates representing the underlying *bla*$_{\text{TEM}}$ genetic background. The workflow comprised multinomial propensity-score estimation, assessment of promoter-group overlap, and regression-based counterfactual prediction for estimating pairwise Average Treatment Effects (ATEs), followed by bootstrap resampling to assess the stability of the estimated effects and permutation testing to determine whether the observed estimates were larger than those obtained under random permutation of the antibiotic phenotypes. All analyses were performed in Python 3.12 using pandas, NumPy and scikit-learn (73–75).

Definition of treatment, outcome, and adjustment covariates and assumptions

Causal inference requires explicit specification of the treatment, outcome, and covariates that define the causal question of interest (76,77). The treatment represents the intervention or exposure whose causal effect is to be estimated, the outcome represents the response variable of interest under the treatment. Covariates are observed characteristics or variables included in the analysis to account for differences among individuals or units that may be relevant to the treatment–outcome relationship (77,78) (Fig. 3).

The treatment variable was represented as:

$$X_i \in \{P_1, P_2, \dots, P_K\},$$

where X_i denotes the observed promoter category assigned to isolate i, and P_k represents the k-th promoter category.

The outcome variable was defined as:

$$Y_i \in \{0,1\},$$

where $Y_i = 1$ represents a resistant phenotype and $Y_i = 0$ represents a susceptible phenotype for the corresponding antibiotic.

The *bla*$_{\text{TEM}}$ variant covariates were represented as:

$$T_i = (t_{i1}, t_{i2}, \dots, t_{iq}),$$

where t_{ij} denotes the presence (1) or absence (0) of the j-th *bla*$_{\text{TEM}}$ variant in isolate i, and q represents the total number of *bla*$_{\text{TEM}}$ variant features included as adjustment covariates.

In the present study, promoter type was defined as the treatment variable (X), antibiotic resistance phenotype as the binary outcome (Y), and the presence or absence of individual *bla*$_{\text{TEM}}$ variants as observed adjustment covariates (T). *bla*$_{\text{TEM}}$ variants were included because differences in the underlying β-lactamase genetic background may contribute to variation in the observed resistance phenotype, and were therefore retained as adjustment covariates when estimating promoter-associated counterfactual effects.

The counterfactual analysis was formulated under the standard causal inference framework of consistency, positivity, and adjustment for observed covariates (77). Consistency requires that each promoter category used in the causal analysis corresponds to a clearly defined promoter motif combination, so that an isolate classified under a given promoter type represents the same

treatment condition throughout the analysis, while positivity refers to sufficient overlap in promoter probabilities across the observed *bla*$_{\text{TEM}}$ variant profiles to support comparisons between promoter types. The *bla*$_{\text{TEM}}$ variant profile was included as the set of observed covariates in the analysis, allowing the promoter effects to be evaluated while accounting for the observed genetic background of each isolate. Other biological or clinical factors that may contribute to resistance were not included as adjustment covariates because of insufficient details. The resulting counterfactual estimates therefore quantify differences in predicted resistance probability associated with changes in promoter type while accounting for the observed *bla*$_{\text{TEM}}$ variant profile.

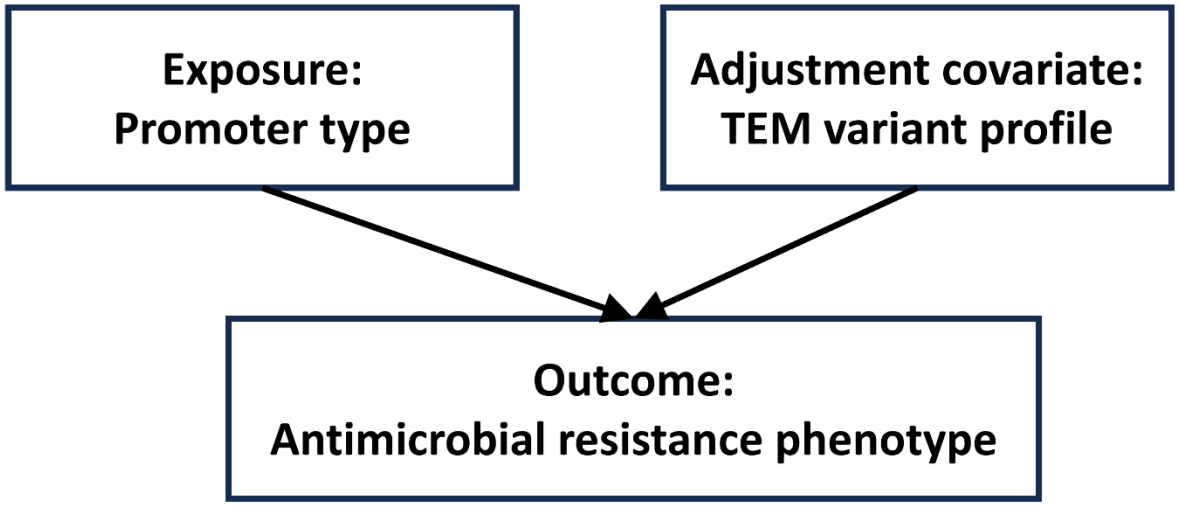


Fig. 3 Directed acyclic graph representing the causal structure specified for the counterfactual analysis. Promoter type was defined as the exposure and antimicrobial resistance phenotype as the outcome. The *bla*$_{\text{TEM}}$ variant profile was included as an adjustment covariate because *bla*$_{\text{TEM}}$ variants may independently influence resistance, allowing counterfactual promoter comparisons to be estimated while holding the underlying *bla*$_{\text{TEM}}$ genetic background constant.

Multinomial propensity score estimation

The first step of the analysis was the estimation of multinomial propensity scores. In causal inference, a propensity score represents the conditional probability of receiving a particular treatment level given the observed covariates (79). Since promoter type consisted of multiple categorical treatment levels, multinomial logistic regression was used to estimate the probability of belonging to each promoter category (80), with the binary presence or absence of individual *bla*$_{\text{TEM}}$ variants included as observed covariates. For each isolate, the fitted model therefore estimated a probability distribution across all promoter categories conditional on its observed *bla*$_{\text{TEM}}$ variant profile, with the probabilities summing to one.

The propensity score for isolate i corresponding to promoter category k was defined as

$$e_i(k) = P(X_i = k \mid T_i) \quad (1)$$

where X_i denotes the promoter type of isolate i, T_i represents the observed *bla*$_{\text{TEM}}$ variant profile, and $e_i(k)$ corresponds to the estimated probability of belonging to promoter category

k conditional on the observed *bla*$_{\text{TEM}}$ variant profile. Thus, each isolate received a set of category-specific propensity scores representing its estimated probability of belonging to each promoter category based on its observed *bla*$_{\text{TEM}}$ genetic background. These estimated propensity scores were subsequently used to evaluate the overlap between promoter categories and assess the positivity assumption before the counterfactual effect estimation.

Assessment of the positivity (overlap) assumption

Following multinomial propensity score estimation, the positivity (overlap) assumption was assessed prior to counterfactual effect estimation. Positivity requires that observations with similar observed covariate profiles have a non-zero probability of belonging to each treatment level under comparison (77). In the present study, pairwise overlap between promoter categories was assessed using the category-specific propensity probabilities estimated from the multinomial logistic regression model. For each pair of promoter categories, an isolate was considered to contribute to shared support when the estimated propensity probability for both promoter categories exceeded 0.01. The overlap fraction was calculated as the proportion of isolates satisfying this criterion:

$$Overlap\ Fraction_{(a,b)} = \frac{\sum_{i=1}^{N} \quad I[e_i(a) > 0.01\ and\ e_i(b) > 0.01]}{N} \qquad (2)$$

where $e_i(a)$ and $e_i(b)$ denote the estimated propensity probabilities for promoter categories a and b, respectively, for isolate i, and N represents the total number of isolates included in the comparison. The threshold of 0.01 was selected as a study-specific criterion to accommodate the sparsity and imbalance of the biological dataset when identifying shared support between promoter categories.

The overlap fraction represents the proportion of isolates with shared propensity support between two promoter categories and serves as a quantitative measure of common support. Larger overlap fractions indicate that a greater proportion of isolates have support for both promoter categories based on their observed *bla*$_{\text{TEM}}$ variant profiles, whereas smaller overlap fractions indicate that fewer isolates have such shared support within the available dataset.

Average treatment effect (ATE) estimation

Following the assessment of positivity, promoter-associated counterfactual effects were estimated using the Average Treatment Effect (ATE). The ATE represents the average difference between the potential outcomes under two alternative treatment conditions in the same analysis population (77,78). In the present study, the ATE was used to quantify the average difference in predicted antimicrobial resistance probability between two promoter categories being compared. For each antibiotic, a separate binary logistic regression model was fitted using antibiotic resistance phenotype as the response variable, promoter type as the treatment variable, and the presence or absence of individual *bla*$_{\text{TEM}}$ variants as adjustment covariates. Promoter categories were one-hot encoded, with one category serving as the reference level to avoid perfect multicollinearity. The fitted model was then used to estimate resistance probabilities under each promoter category included in the pairwise comparison.

For each isolate and promoter comparison, two predicted resistance probabilities were generated using the fitted logistic regression model. The promoter category was first set to p_a and then to p_b, while the observed *bla*$_{\text{TEM}}$ variant profile of the isolate remained unchanged in both predictions. The resulting probabilities were denoted as $\hat{\pi}_i(p_a)$ and $\hat{\pi}_i(p_b)$, where $\hat{\pi}_i(p)$

represents the model-predicted probability of resistance for isolate i under promoter category p. The pairwise ATE was calculated by averaging the difference between these predicted probabilities across all isolates included in the corresponding antibiotic-specific analysis:

$$ATE_{p_a \to p_b} = \frac{1}{N}\sum_{i=1}^{N} [\hat{\pi}_i(p_b) - \hat{\pi}_i(p_a)] \quad (3)$$

where N denotes the total number of isolates included in the analysis for the corresponding antibiotic. Thus, the ATE represents the average model-predicted change in resistance probability associated with evaluating promoter category p_b instead of p_a, while retaining the observed *bla*$_{\text{TEM}}$ genetic background of each isolate.

A positive ATE indicates that the predicted resistance probability was higher under promoter category p_b than under p_a, whereas a negative ATE indicates a lower predicted resistance probability under p_b. For example, an ATE of 0.20 for the transition $p_a \to p_b$ represents an average increase of 20 percentage points in the predicted resistance probability under p_b.

Bootstrap resampling for effect stability

Following ATE estimation, bootstrap resampling was used to assess the stability of the estimated counterfactual effects across repeated sampling of the observed dataset. A total of 1,000 bootstrap samples were generated by sampling isolates from the complete dataset with replacement, with each sample containing the same total number of observations as the original dataset. For each bootstrap sample, ATEs were recalculated separately for each antibiotic using isolates with available resistance phenotype and promoter information.

ATE stability for each antibiotic-specific promoter comparison was calculated across the 1,000 bootstrap iterations. An absolute ATE threshold of 0.01 was used as a study-specific criterion to exclude near-zero bootstrap estimates, given the sparsity and imbalance of the promoter-phenotype data. A promoter comparison was considered stable when at least 80% of the bootstrap iterations produced an ATE exceeding 0.01 in absolute magnitude and retaining the same direction as the ATE estimated from the original dataset.

ATE stability was calculated as:

$$\frac{No.\,of\ bootstrap\ iterations\ satisfying\ both\ magnitude\ \&\ direction\ criteria}{Total\ number\ of\ bootstrap\ iterations}$$

Permutation testing of ATE significance

Following the bootstrap assessment of ATE stability, permutation testing was performed to evaluate whether the magnitude of the observed ATE was greater than that expected under a null distribution generated by randomizing the antibiotic resistance phenotypes. For each antibiotic, the available resistance phenotype labels were randomly shuffled across isolates in 500 permutation iterations, while the observed promoter categories and *bla*$_{\text{TEM}}$ variant profiles were retained unchanged. The ATE estimation procedure was repeated for each permuted dataset to generate a null distribution for every antibiotic-specific promoter comparison.

For each comparison, a two-sided empirical permutation p-value was calculated as the proportion of permuted ATE estimates whose absolute magnitude was equal to or greater than that of the observed ATE:

$$p_{perm} = \frac{\#(|ATE_{perm}| \geq |ATE_{obs}|)}{N_{perm}} \quad (4)$$

where ATE_{obs} denotes the ATE estimated from the original dataset, ATE_{perm} denotes the ATE obtained from a permuted dataset, and $N_{perm} = 500$. Comparisons with $p_{perm} \leq 0.05$ were considered to indicate that the observed promoter-associated effect was larger in absolute magnitude than the effects obtained under the phenotype-randomized null distribution.

Following bootstrap resampling and permutation testing, promoter comparisons were evaluated using the predefined criteria of ATE stability ≥ 0.80 and permutation p-value ≤ 0.05. Promoter comparisons satisfying both criteria were considered to show stable promoter-associated counterfactual effects with permutation-based statistical support for the corresponding antibiotic resistance phenotype. The stability criterion required that at least 80% of the bootstrap iterations produced an ATE with an absolute magnitude greater than 0.01 and the same direction as the ATE estimated from the original dataset, whereas the permutation test assessed whether the magnitude of the observed ATE was unusually large relative to the null distribution generated by randomizing the antibiotic resistance phenotypes. These complementary criteria were used jointly to identify promoter comparisons whose ATEs remained above the predefined magnitude threshold and directionally consistent across bootstrap resampling, while also showing an observed effect greater than that generally obtained under the phenotype-randomized null distribution, providing a basis for subsequent hypothesis generation.

Computational assessment of resistance phenotypes associated with putative novel promoter motifs

To further assess the identified putative novel promoter motifs in relation to β-lactam resistance phenotypes, isolates carrying these motifs were subjected to additional computational screening. Novel-motif-containing isolates were first separated from the analyzed dataset and subsequently filtered to retain isolates carrying a single *bla*$_{\text{TEM}}$ variant without any additional β-lactamase genes, thereby reducing the potential contribution of other β-lactamases to the observed resistance phenotype. Resistance phenotype information was then examined for the resulting isolates, yielding 12 isolates with available phenotypic data. Analysis of these isolates identified five putative novel motif configurations with variations in the -10 and/or -35 regions. Of these, three motifs that were observed in more than one isolate were selected for subsequent structure-based assessment using molecular dynamics simulation.

Structure-based assessment of putative novel promoter motifs using molecular dynamics simulation

To assess whether the identified putative novel promoter motifs retained structural interactions with σ^{70} comparable to those of the canonical promoter motifs, a structure-based analysis followed by molecular dynamics (MD) simulation was performed. The DNA-bound cryo-EM structure of the *E. coli* RNA polymerase σ^{70}-containing promoter open complex (PDB ID: 6CA0) (81) was used as the structural template. Domain 2 (residues 360-465) and Domain 4 (residues 530-612) of the σ^{70} (RpoD) subunit, corresponding to the regions interacting with the -10 and -35 promoter elements, respectively, were extracted for subsequent structural analysis. The canonical and novel -10 and -35 motifs were generated within the corresponding DNA

regions of the structural template by nucleotide substitution, while retaining the original σ^{70} structure. An 11-bp double-stranded DNA segment was used to represent the -35 region, whereas a 12-nucleotide single-stranded DNA segment was used to represent the -10 region, retaining the local DNA context required for analysis of σ^{70}-promoter interactions. Both canonical and novel -10 and -35 motif-containing systems were subjected to MD simulation using the same simulation protocol. Prior to simulation, the complexes were pre-processed with terminal capping enabled. Each system was subsequently solvated using the TIP3P water model in an orthorhombic simulation box with a 10 Å buffer distance around the solute. The systems were neutralized with appropriate counterions and supplemented with NaCl to a final concentration of 0.15 M. Molecular dynamics simulations were performed for 100 ns using the OPLS4 force field in Desmond v8.3 (Schrödinger Release 2025-3) (82,83) at a temperature of 300 K and a pressure of 1.01325 bar. Trajectory frames were recorded at 100-ps intervals and subsequently used for protein-DNA interaction analysis. Protein-DNA interactions across the simulation trajectory were calculated using the Python-based interaction analysis script available in Schrödinger Release 2025-3.

Acknowledgements

The authors thank the Department of Biotechnology, Government of India (grant no. BT/PR40144/BTIS/137/46/2022, BT/PR40150/BTIS/137/81/2023) for the funding. The author, SR, thanks the Department of Biotechnology for the manpower support.

Author contribution

SM & SR, conceptualization, study design. SAC data download and analysis. SR interpretation and script execution. SR & SM Draft and revision of the manuscript. All authors have read and agreed to the final version of the manuscript.

Funding

This work was supported by a grant from the Department of Biotechnology, Government of India (grant no BT/PR40144/BTIS/137/46/2022, BT/PR40150/BTIS/137/81/2023)

Data availability

The study used publicly available data from the BVBRC. The details of the dataset are provided in the supplementary files. The code used for causal analysis and for figure generation is submitted in the GitHub repository:

https://github.com/SrimathyRamachandran/Counterfactual-causal-analysis

Supplementary Material 1. Contains Tables S1-S5.

Supplementary Material 2. Contains Figs S1-S6.

Declarations

Declarations of AI-assisted technologies in writing process

While preparing this work, the authors used ChatGPT to improve the readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

1. Blair JMA, Webber MA, Baylay AJ, Ogbolu DO, Piddock LJ V. Molecular mechanisms of antibiotic resistance. Nat Rev Microbiol. 2015 Jan;13(1):42–51.
2. Davies J, Davies D. Origins and evolution of antibiotic resistance. Microbiol Mol Biol Rev. 2010 Sep;74(3):417–33.
3. Browning DF, Busby SJ. The regulation of bacterial transcription initiation. Nat Rev Microbiol. 2004 Jan;2(1):57–65.
4. Feklístov A, Sharon BD, Darst SA, Gross CA. Bacterial sigma factors: a historical, structural, and genomic perspective. Annu Rev Microbiol. 2014;68:357–76.
5. Walsh C. Molecular mechanisms that confer antibacterial drug resistance. Nature [Internet]. 2000;406(6797):775–81. Available from: https://doi.org/10.1038/35021219
6. Bush K, Bradford PA. Interplay between β-lactamases and new β-lactamase inhibitors. Nat Rev Microbiol [Internet]. 2019;17(5):295–306. Available from: https://doi.org/10.1038/s41579-019-0159-8
7. Ambler RP. The structure of beta-lactamases. Philos Trans R Soc London Ser B, Biol Sci. 1980 May;289(1036):321–31.
8. Bush K, Jacoby GA. Updated functional classification of beta-lactamases. Antimicrob Agents Chemother. 2010 Mar;54(3):969–76.
9. A. BP. Extended-Spectrum β-Lactamases in the 21st Century: Characterization, Epidemiology, and Detection of This Important Resistance Threat. Clin Microbiol Rev [Internet]. 2001 Oct 1;14(4):933–51. Available from: https://doi.org/10.1128/cmr.14.4.933-951.2001
10. Palzkill T. Structural and Mechanistic Basis for Extended-Spectrum Drug-Resistance Mutations in Altering the Specificity of TEM, CTX-M, and KPC β-lactamases. Front Mol Biosci [Internet]. 2018;Volume 5-. Available from: https://www.frontiersin.org/journals/molecular-biosciences/articles/10.3389/fmolb.2018.00016
11. Cheng Q, Cheung YC, Chan EWC, Wong KY, Chen S. Unveiling the evolution routes of TEM-type extended-spectrum β-lactamases. Int J Antimicrob Agents. 2022 Jan;59(1):106498.
12. Salverda MLM, De Visser JAGM, Barlow M. Natural evolution of TEM-1 β-lactamase: experimental reconstruction and clinical relevance. FEMS Microbiol Rev. 2010 Nov;34(6):1015–36.

13. Shaikh S, Fatima J, Shakil S, Rizvi SMD, Kamal MA. Antibiotic resistance and extended spectrum beta-lactamases: Types, epidemiology and treatment. Saudi J Biol Sci. 2015 Jan;22(1):90–101.

14. Castanheira M, Simner PJ, Bradford PA. Extended-spectrum β-lactamases: an update on their characteristics, epidemiology and detection. JAC-Antimicrobial Resist [Internet]. 2021 Sep 1;3(3):dlab092. Available from: https://doi.org/10.1093/jacamr/dlab092

15. Carattoli A. Resistance Plasmid Families in *Enterobacteriaceae*. Antimicrob Agents Chemother [Internet]. 2009;53(6):2227–38. Available from: https://journals.asm.org/doi/abs/10.1128/aac.01707-08

16. Partridge SR, Hall RM. Evolution of Transposons Containing *bla*$_{TEM}$ Genes. Antimicrob Agents Chemother [Internet]. 2005;49(3):1267–8. Available from: https://journals.asm.org/doi/abs/10.1128/aac.49.3.1267-1268.2005

17. Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile Genetic Elements Associated with Antimicrobial Resistance. Clin Microbiol Rev. 2018 Oct;31(4).

18. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B, Guerra B, et al. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. J Antimicrob Chemother. 2018 May;73(5):1121–37.

19. Lartigue MF, Leflon-Guibout V, Poirel L, Nordmann P, Nicolas-Chanoine MH. Promoters P3, Pa/Pb, P4, and P5 upstream from bla(TEM) genes and their relationship to beta-lactam resistance. Antimicrob Agents Chemother. 2002 Dec;46(12):4035–7.

20. Zhou K, Tao Y, Han L, Ni Y, Sun J. Piperacillin-Tazobactam (TZP) Resistance in Escherichia coli Due to Hyperproduction of TEM-1 β-Lactamase Mediated by the Promoter Pa/Pb. Front Microbiol. 2019;10:833.

21. Wang Y, Batra A, Schulenburg H, Dagan T. Gene sharing among plasmids and chromosomes reveals barriers for antibiotic resistance gene transfer. Philos Trans R Soc London Ser B, Biol Sci. 2022 Jan;377(1842):20200467.

22. Wang X, Zhang H, Yu S, Li D, Gillings MR, Ren H, et al. Inter-plasmid transfer of antibiotic resistance genes accelerates antibiotic resistance in bacterial pathogens. ISME J [Internet]. 2024 Jan 1;18(1):wrad032. Available from: https://doi.org/10.1093/ismejo/wrad032

23. Hernandez-Beltran JCR, Rodríguez-Beltrán J, Aguilar-Luviano OB, Velez-Santiago J, Mondragón-Palomino O, MacLean RC, et al. Plasmid-mediated phenotypic noise leads to transient antibiotic resistance in bacteria. Nat Commun. 2024 Mar;15(1):2610.

24. Botelho J, Schulenburg H. The Role of Integrative and Conjugative Elements in Antibiotic Resistance Evolution. Trends Microbiol. 2021 Jan;29(1):8–18.

25. J. HC, M. HR. IS26 and the IS26 family: versatile resistance gene movers and genome reorganizers. Microbiol Mol Biol Rev [Internet]. 2024 Mar 4;88(2):e00119-22. Available from: https://doi.org/10.1128/mmbr.00119-22

26. Chen Y, Tenover FC, Koehler TM. Beta-lactamase gene expression in a penicillin-resistant Bacillus anthracis strain. Antimicrob Agents Chemother. 2004 Dec;48(12):4873–7.

27. Liu P, Xue H, Wu Z, Ma J, Zhao X. Effect of bla regulators on the susceptible phenotype and phenotypic conversion for oxacillin-susceptible mecA-positive staphylococcal isolates. J Antimicrob Chemother. 2016 Aug;71(8):2105–12.

28. Romero-Muñoz M, Pulido MR, Recacha E, Díaz-Díaz S, Murillo-Torres M, Docobo-Pérez F, et al. Implications of gene expression heterogeneity in the interplay between acquired resistance and bacterial metabolism. Sci Rep [Internet]. 2025;15(1):26632. Available from: https://doi.org/10.1038/s41598-025-10760-1

29. Ramachandran S, K D, M S, R MR, S A, V A, et al. Unveiling meropenem resistance and co-resistance patterns in Klebsiella pneumoniae and Acinetobacter baumannii: a global genome analysis using ML/DL and association mining. BMC Microbiol [Internet]. 2025;25(1):699. Available from: https://doi.org/10.1186/s12866-025-04394-6

30. Wan F, Tong W, Wu W, Liu J, Huang H, Zhang P, et al. Utilizing whole genome sequencing data for machine learning driven prediction of antibiotic resistance in Escherichia coli. Front Microbiol [Internet]. 2026;Volume 17. Available from: https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2026.1842717

31. Sherry NL, Lee JYH, Giulieri SG, Connor CH, Horan K, Lacey JA, et al. Genomics for antimicrobial resistance—progress and future directions. Antimicrob Agents Chemother [Internet]. 2025;69(5):e01082-24. Available from: https://journals.asm.org/doi/abs/10.1128/aac.01082-24

32. Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EPC, de la Cruz F. Mobility of plasmids. Microbiol Mol Biol Rev. 2010 Sep;74(3):434–52.

33. Siguier P, Gourbeyre E, Chandler M. Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol Rev [Internet]. 2014 Sep 1;38(5):865–91. Available from: https://doi.org/10.1111/1574-6976.12067

34. Rocha EPC. The organization of the bacterial genome. Annu Rev Genet. 2008;42:211–33.

35. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, et al. Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. PLoS Genet. 2009 Jan;5(1):e1000344.

36. Ortega A, Oteo J, Aranzamendi-Zaldumbide M, Bartolomé RM, Bou G, Cercenado E, et al. Spanish Multicenter Study of the Epidemiology and Mechanisms of Amoxicillin-Clavulanate Resistance in Escherichia coli. Antimicrob Agents Chemother [Internet]. 2012;56(7):3576–81. Available from: https://journals.asm.org/doi/abs/10.1128/aac.06393-11

37. Singh NS, Singhal N, Virdi JS. Genetic Environment of bla(TEM-1), bla(CTX-M-15), bla(CMY-42) and Characterization of Integrons of Escherichia coli Isolated From an Indian Urban Aquatic Environment. Front Microbiol. 2018;9:382.

38. Sutcliffe JG. Complete nucleotide sequence of the Escherichia coli plasmid pBR322. Cold Spring Harb Symp Quant Biol. 1979;43 Pt 1:77–90.

39. Goussard S, Courvalin P. Updated Sequence Information for TEM β-Lactamase Genes. Antimicrob Agents Chemother [Internet]. 1999;43(2):367–70. Available from:

https://journals.asm.org/doi/abs/10.1128/aac.43.2.367

40. San Millan A, Escudero JA, Gifford DR, Mazel D, MacLean RC. Multicopy plasmids potentiate the evolution of antibiotic resistance in bacteria. Nat Ecol Evol. 2016 Nov;1(1):10.

41. Fournier B, Gravel A, Hooper DC, Roy PH. Strength and regulation of the different promoters for chromosomal beta-lactamases of Klebsiella oxytoca. Antimicrob Agents Chemother. 1999 Apr;43(4):850–5.

42. Sutcliffe JG. Nucleotide sequence of the ampicillin resistance gene of Escherichia coli plasmid pBR322. Proc Natl Acad Sci U S A. 1978 Aug;75(8):3737–41.

43. Matlock W, Rodger G, Pritchard E, Colpus M, Kapel N, Barrett L, et al. Escherichia coli phylogeny drives co-amoxiclav resistance through variable expression of TEM-1 beta-lactamase. Nat Commun [Internet]. 2025;16(1):8669. Available from: https://doi.org/10.1038/s41467-025-63714-6

44. Wu PJ, Shannon K, Phillips I. Effect of hyperproduction of TEM-1 beta-lactamase on in vitro susceptibility of Escherichia coli to beta-lactam antibiotics. Antimicrob Agents Chemother. 1994 Mar;38(3):494–8.

45. Wu PJ, Shannon K, Phillips I. Mechanisms of hyperproduction of TEM-1 beta-lactamase by clinical isolates of Escherichia coli. J Antimicrob Chemother. 1995 Dec;36(6):927–39.

46. Zhou XY, Bordon F, Sirot D, Kitzis MD, Gutmann L. Emergence of clinical isolates of Escherichia coli producing TEM-1 derivatives or an OXA-1 beta-lactamase conferring resistance to beta-lactamase inhibitors. Antimicrob Agents Chemother. 1994 May;38(5):1085–9.

47. Poyart C, Mugnier P, Quesne G, Berche P, Trieu-Cuot P. A novel extended-spectrum TEM-type beta-lactamase (TEM-52) associated with decreased susceptibility to moxalactam in Klebsiella pneumoniae. Antimicrob Agents Chemother. 1998 Jan;42(1):108–13.

48. Hubbard ATM, Mason J, Roberts P, Parry CM, Corless C, van Aartsen J, et al. Piperacillin/tazobactam resistance in a clinical isolate of Escherichia coli due to IS26-mediated amplification of blaTEM-1B. Nat Commun [Internet]. 2020;11(1):4915. Available from: https://doi.org/10.1038/s41467-020-18668-2

49. Mbanga J, Amoako DG, Abia ALK, Allam M, Ismail A, Essack SY. Genomic Insights of Multidrug-Resistant Escherichia coli From Wastewater Sources and Their Association With Clinical Pathogens in South Africa. Front Vet Sci [Internet]. 2021;Volume 8-. Available from: https://www.frontiersin.org/journals/veterinary-science/articles/10.3389/fvets.2021.636715

50. Goussard S, Sougakoff W, Mabilat C, Bauernfeind A, Courvalin P. An ISI-like element is responsible for high-level synthesis of extended-spectrum β-lactamase TEM-6 in Enterobacteriaceae. Microbiology [Internet]. 1991;137(12):2681–7. Available from: https://www.microbiologyresearch.org/content/journal/micro/10.1099/00221287-137-12-2681

51. Kanjilal S, Jeong H, Ma Y, Wei A, Yang K, Sontag D. 2238. Estimating the impact of

antibiotic exposure on antibiotic resistance in uncomplicated UTI using machine learning causal inference. Vol. 9, Open Forum Infectious Diseases. 2022.

52. Lim C, Pouwels KB, Lee SJ, Mo Y, van Doorn RH, Turner P, et al. The burden of antimicrobial-resistant bacterial infections: a causal perspective. Nat Commun [Internet]. 2026;17(1):5847. Available from: https://doi.org/10.1038/s41467-026-74934-9

53. Gago J, Boyer C, Lipsitch M. How Should We Study the Indirect Effects of Antimicrobial Treatment Strategies?: A Causal Perspective. Epidemiology. 2025;37(1):88–97.

54. Campbell EA, Muzzin O, Chlenov M, Sun JL, Olson CA, Weinman O, et al. Structure of the bacterial RNA polymerase promoter specificity sigma subunit. Mol Cell. 2002 Mar;9(3):527–39.

55. Patikoglou GA, Westblade LF, Campbell EA, Lamour V, Lane WJ, Darst SA. Crystal structure of the Escherichia coli regulator of sigma70, Rsd, in complex with sigma70 domain 4. J Mol Biol. 2007 Sep;372(3):649–59.

56. Klocko AD, Wassarman KM. 6S RNA binding to Esigma(70) requires a positively charged surface of sigma(70) region 4.2. Mol Microbiol. 2009 Jul;73(2):152–64.

57. Gardella T, Moyle H, Susskind MM. A mutant Escherichia coli σ70 subunit of RNA polymerase with altered promoter specificity. J Mol Biol [Internet]. 1989;206(4):579–90. Available from: https://www.sciencedirect.com/science/article/pii/0022283689905676

58. Ross W, Schneider DA, Paul BJ, Mertens A, Gourse RL. An intersubunit contact stimulating transcription initiation by E coli RNA polymerase: interaction of the alpha C-terminal domain and sigma region 4. Genes Dev. 2003 May;17(10):1293–307.

59. Chen H, Tang H, Ebright RH. Functional interaction between RNA polymerase alpha subunit C-terminal domain and sigma70 in UP-element- and activator-dependent transcription. Mol Cell. 2003 Jun;11(6):1621–33.

60. Rhodius VA, Busby SJ. Interactions between activating region 3 of the Escherichia coli cyclic AMP receptor protein and region 4 of the RNA polymerase sigma(70) subunit: application of suppression genetics. J Mol Biol. 2000 Jun;299(2):311–24.

61. Zhang Y, Feng Y, Chatterjee S, Tuske S, Ho MX, Arnold E, et al. Structural basis of transcription initiation. Science. 2012 Nov;338(6110):1076–80.

62. Fenton MS, Lee SJ, Gralla JD. Escherichia coli promoter opening and -10 recognition: mutational analysis of sigma70. EMBO J. 2000 Mar;19(5):1130–7.

63. Panaghie G, Aiyar SE, Bobb KL, Hayward RS, de Haseth PL. Aromatic amino acids in region 2.3 of Escherichia coli sigma 70 participate collectively in the formation of an RNA polymerase-promoter open complex. J Mol Biol. 2000 Jun;299(5):1217–30.

64. Schroeder LA, Karpen ME, deHaseth PL. Threonine 429 of Escherichia coli σ70 is a Key Participant in Promoter DNA Melting by RNA Polymerase. J Mol Biol [Internet]. 2008;376(1):153–65. Available from: https://www.sciencedirect.com/science/article/pii/S0022283607015690

65. Schroeder LA, Gries TJ, Saecker RM, Record MT, Harris ME, deHaseth PL. Evidence

for a Tyrosine–Adenine Stacking Interaction and for a Short-lived Open Intermediate Subsequent to Initial Binding of Escherichia coli RNA Polymerase to Promoter DNA. J Mol Biol [Internet]. 2009;385(2):339–49. Available from: https://www.sciencedirect.com/science/article/pii/S0022283608013016

66. Callaci S, Heyduk T. Conformation and DNA Binding Properties of a Single-Stranded DNA Binding Region of σ70 Subunit from Escherichia coli RNA Polymerase Are Modulated by an Interaction with the Core Enzyme. Biochemistry [Internet]. 1998 Feb 14;37(10):3312–20. Available from: https://doi.org/10.1021/bi972041m

67. Feklistov A, Darst SA. Structural basis for promoter-10 element recognition by the bacterial RNA polymerase σ subunit. Cell. 2011 Dec;147(6):1257–69.

68. Waldburger C, Gardella T, Wong R, Susskind MM. Changes in conserved region 2 of Escherichia coli sigma 70 affecting promoter recognition. J Mol Biol. 1990 Sep;215(2):267–76.

69. Olson RD, Assaf R, Brettin T, Conrad N, Cucinell C, Davis JJ, et al. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. Nucleic Acids Res [Internet]. 2023 Jan 6;51(D1):D678–89. Available from: https://doi.org/10.1093/nar/gkac1003

70. Seemann T. Abricate: Mass Screening of Contigs for Antimicrobial and Virulence Genes [Internet]. Available from: https://github.com/tseemann/abricate

71. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res. 2023 Jan;51(D1):D690–9.

72. Xie Z, Tang H. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. Bioinformatics [Internet]. 2017 Nov 1;33(21):3340–7. Available from: https://doi.org/10.1093/bioinformatics/btx433

73. McKinney W. Data Structures for Statistical Computing in Python. SciPy 2010 [Internet]. 2010; Available from: https://doi.org/10.25080/Majora-92bf1922-00a

74. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature [Internet]. 2020;585(7825):357–62. Available from: https://doi.org/10.1038/s41586-020-2649-2

75. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res [Internet]. 2011;12(85):2825–30. Available from: http://jmlr.org/papers/v12/pedregosa11a.html

76. Pearl J. Causality. 2nd ed. Cambridge University Press; 2009.

77. Hernan, M.A., & Robins JM. Causal Inference: What If (1st ed.). CRC Press. 2023;

78. Imbens GW, Rubin DB. Causal inference: For statistics, social, and biomedical sciences an introduction. Causal Inference: For Statistics, Social, and Biomedical Sciences an Introduction. Cambridge University Press; 2015. 1–625 p.

79. ROSENBAUM PR, RUBIN DB. The central role of the propensity score in observational studies for causal effects. Biometrika [Internet]. 1983 Apr 1;70(1):41–55. Available from: https://doi.org/10.1093/biomet/70.1.41

80. Imbens GW. The role of the propensity score in estimating dose-response functions. Biometrika [Internet]. 2000 Sep 1;87(3):706–10. Available from: https://doi.org/10.1093/biomet/87.3.706

81. Narayanan A, Vago FS, Li K, Qayyum MZ, Yernool D, Jiang W, et al. Cryo-EM structure of Escherichia coli σ(70) RNA polymerase and promoter DNA complex revealed a role of σ non-conserved region during the open complex formation. J Biol Chem. 2018 May;293(19):7367–75.

82. Research DES, Schrödinger LLC. Schrödinger Release 2025-3: Desmond Molecular Dynamics System; Maestro-Desmond Interoperability Tools. New York, NY: D. E. Shaw Research; Schrödinger, LLC; 2025.

83. Lu C, Wu C, Ghoreishi D, Chen W, Wang L, Damm W, et al. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. J Chem Theory Comput [Internet]. 2021 Jun 7;17(7):4291–300. Available from: https://doi.org/10.1021/acs.jctc.1c00302.

Supplementary Files

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

- SuplementaryFile2.pdf
- SupplementaryFile1.xlsx