**Declaration:**

I declare that I have not employed a Chat-GPT-3.5 to assist in creating this report.

**Comparative genomic and structural analysis of antimicrobial resistance genes across *Escherichia coli/Shigella* strains.**

**Abstract**

Antimicrobial resistance (AMR) gene evolution poses a global health emergency, with *Escherichia coli/Shigella* strains acting as a key reservoir. This study analysed the sequences of three clinical isolates (GN3, GN6, and GN9) with the E. coli K12 as a reference genome. Pangenome analysis with the Resistance Gene Identifier (RGI) tool revealed a conserved set of 135 AMR genes, with EF-Tu and efflux pump families (RND, ABC, MFS) most prevalent. Mutations were detected in *acrR* and associated with efflux pump regulation and multidrug resistance.

Two paralogs of the AcrR transcriptional regulator were analysed. AcrR1 contained two C-terminal missense mutations, and AcrR2 had seven missense mutations. While domains were consistent, AcrR2 exhibited extensive C-terminal mutations and structural divergence. The C-terminal domain was identified as the ligand-binding domain, associated with regulation, post-transcriptional modifications, and antibiotic binding. Furthermore, evidence for a duplication event with sub-functionalisation is supported by distinct paralog clades and substantial structural deviation.

**Introduction**

With resistance to all current clinical antibiotics detected and only a few novel drugs in development, antimicrobial resistance (AMR) is a global health emergency (Darby et al., 2022). Substantial progress has been made in understanding antibiotic functionality and bacterial antimicrobial mechanisms (Figure 1; Darby et al., 2022).

As a primary reservoir for the development and spread of antimicrobial resistance genes (ARGs), *E. coli* is crucial in understanding the mechanisms underlying resistance to aid in novel treatment design (Poirel et al., 2018).

Although *Shigella* is a four-species genus, evidence suggests it represents a specialised lineage of the highly diverse species of *E. coli.* The Genome Taxonomy Database (GTDB) reclassified *Shigella* species as Escherichia species and nearly 80% of *E. coli* strains were reclassified to new species. For example, *E. coli* K-12 changed to *E. flexneri* as it wascloser related to *Shigella flexneri* than the *E. coli* type strain (Parks et al., 2021). Several lineages of *Shigella* strains were derived via the independent acquisition of pINV virulence plasmid, which is vital for host invasion and intracellular survival (Lan and Reeves, 2002). As humans are the only natural hosts of *Shigella,* adaptation to the human environment with frequent proximity to antibiotics likely drives the acquisition and proliferation of ARGs. The shared resistance mechanisms and evolutionary relationship between *E. coli* and *Shigella* make comparative analysis particularly valuable to uncover the genomic and functional characteristics of AMR.

Substantial progress in the field of AMR has revealed major mechanisms bacteria use to resist the inhibitory or fatal effects of antibiotics (Figure 1). Efflux pumps, especially the AcrAB-TolC system, play crucial roles in the active export of a wide range of antibiotics (Figure 1.6). Mutations in the acrR can lead to AcrAB overexpression, enhancing resistance. Therefore, understanding the structural and functional features of acrR mutant variants across *E. coli/Shigella* strains is crucial for mitigating resistance development. This study explores the genetic and functional attributes of AMR pathways to understand the difference in mechanisms and distributions of ARGs between *Shigella* and *E. coli* strains.

*A diagram of a cell

AI-generated content may be incorrect.*

**Figure 1. Molecular mechanisms of antibiotic resistance in bacteria.** Bacteria generally evade antibiotics via six strategies: 1) Porin downregulation to decrease antibiotic influx; 2) proteins shield the target to protect against antibiotic binding; 3) antibiotic inactivation by chemical modification or enzymatic degradation; 4) alternative pathway usage to bypass the target; 5) Modification to the target site through enzymatic alterations or mutation; 6) transmembrane pumps expel antibiotics via active efflux to minimise intracellular concentrations. Based on research from Darby et al., (2022). Created in https://BioRender.com.

**Materials and Methods**

*Sample Collection and DNA Extraction*

Three clinical samples were collected from three separate septic patients identified three strains of *E. coli/Shigella* (GN3, GN6, GN9) sequenced with PacBio HiFi reads at a 30-fold coverage. The bacteria were cultured on standard LB liquid media overnight at 37°C and DNA was extracted with a Qiagen Power Max Soil kit following manufacturer’s instructions (<https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/microbial-dna/dneasy-powermax-soil-kit>).

*Library Preparation and Sequencing*

Libraries were prepared for PacBio sequencing utilising the SMRTbell Prep Kit 3.0, focusing on size selection to optimise read quality and length. Sequencing was conducted on a Sequel IIe machine with SMRT Link v13.0 base-calling, following the manufacturer's settings for HiFi reads.

*Bioinformatics Analysis*

The genomes were assembled using *Flye*, and the quality was assessed using FastQC and Quast. Data was accessed and analysed using the Windows Subsystem for Linux (WSL), Ubuntu.

The *Escherichia coli/Shigella* and referencegenomes in GFF file format were loaded into Panaroo (<https://github.com/gtonkinhill/panaroo>) for core genome alignment. All sequences were uploaded to the resistance gene identifier (RGI) 6.0.2 tool with default parameters to ensure thorough and consistent analysis (Sweta Padma Routray et al., 2024). The resistant gene families and mechanisms were compared across strains using R Studio. Using diamond-blastx (https://github.com/bbuchfink/diamond) FASTA files were used to perform homology searches against the Virulence Factor Database (VFDB) (Liu et al., 2018). An E-value cutoff of 0.05 and a minimum identity threshold of 80% were used to retain high-confidence matches. These virulence factor genes were further annotated with NCBI descriptions using Entrez within Python (version 3.10).

The AMR genes were compared between strains through upset plots and heatmaps to highlight the conservation and variation of these genes. Using R, the top overlapping AMR and VFDB protein-coding sequences were extracted from annotated assemblies and aligned with Clustal Omega to identify strain-specific mutations. Phylogenetic trees were created using MEGA11, using the T92+I model with 500 bootstraps. The sequences were translated for domain analysis using InterPro.

The protein paralogs were modelled using ColabFold v1.5.5, aligned and visualised with PyMOL, with mutations marked as sticks and inter-residue distances calculated. All proteins were aligned with consistent orientation for comparison of structural features, including C-terminal extension.

**Results and Discussion**

*Pangenome Analysis*

Panaroo pangenome analysis uncovered substantial conservation throughout the strains. While this was predominantly the core genome (yellow), there was a smaller proportion of accessory (red) and unique (blue) genes, suggesting strain-specific traits (Figure 2). Overall, there were 4347 core genes, 542 accessory genes, and 809 unique genes. Core genes are highly conserved all analysed strains and are generally essential for cellular functioning, with functions involved in growth, replication and transcription, translation, and metabolic regulation (Gordienko, Kazanov and Gelfand, 2013).

The genomes were hierarchically clustered based on gene content (Figure 3), revealing that GN6 and GN9 share the closest relation, followed by GN3, and the reference (*E. coli K12*) is the least closely related. This pattern was supported by the presence/absence heatmaps (Figures 4 and 5), with GN6 and GN9 displaying similar gene presence and absence patterns. Furthermore, when filtered to non-core accessory genes, the reference genome displayed the least gene presence, suggesting that the three strains share more recent ecological or evolutionary similarities (Figure 5). This genetic variability highlights the roles of accessory and unique genes in adaptive evolution, with the potential acquisition of AMR or virulence factors.

A pie chart with a red blue and yellow triangle

Description automatically generated

**Figure 2. Pangenome composition derived from Panaroo analysis across *Escherichia coli/Shigella* strains**. The pie chart depicts the pangenome composition of three bacterial strains (GN3, GN6, and GN9), utilising *E. coli k12* as a reference genome. The core genome (yellow) signifies genes common across all analysed genomes, indicating conservation of essential functions. The accessory genome (red) represents genes shared by two or more, but not all, genomes. The unique genes (blue) are only present in a single genome. This demonstrates the distribution of these genes.

A diagram of a structure

Description automatically generated

**Figure 3. Hierarchical clustering of *Escherichia coli/Shigella* strain genomes based on gene content similarity.** The dendrogram depicts the relationship between three bacterial strains (GN3, GN6, and GN9), utilising *E. coli k12* as a reference genome. The height of each branch represents the genetic distance between the genomes, with higher values indicating less similarity.

A green and red squares

Description automatically generated

**Figure 4. Gene presence/absence heatmap illustrating binary core and accessory genome clustering across *Escherichia coli/Shigella* strains.** The heatmaps are coloured by gene presence (green) or gene absence (red). Hierarchical clustering of the genomes (columns) and genes (rows) highlights genomic diversity between strains. The gene presence/absence matrix was generated from Panaroo analysis of 3 strains (GN3, GN6, GN9), using *E. coli* K12 as a reference genome.

A screenshot of a graph

Description automatically generated

**Figure 5. Non-core accessory gene presence/absence heatmap illustrating binary genome clustering across *Escherichia coli/Shigella* strains.** Non-core accessory genes are present in multiple, but not all genomes, including GN3, GN6, GN9 and the reference genome *E. coli* K12. The heatmaps are coloured by gene presence (green) or gene absence (red). Hierarchical clustering of the genomes (columns) and genes (rows) highlights genomic diversity between strains.

*AMR Gene Analysis*

The AMR analysis revealed varying conservation and distribution of AMR genes across the four analysed genomes. The Upset plot (Figure 4) depicts the distribution of AMR genes with 135 shared by all genomes, 52 explicit to the *E. coli 12* (reference)genome, 12 shared by reference, GN9 and GN6, and 12 shared by reference and GN9. Figure 7 depicts a heatmap of the AMR gene presence (green) or absence (red), showing most AMR genes are shared across all strains. While *E. coli* K12 contained all identified AMR genes, GN3 clustered individually with few absent genes, and GN6 and GN9 displayed clustering with moderately more absences.

A graph with numbers and a number

Description automatically generated

**Figure 6. The distribution of antimicrobial resistance (AMR) genes across *Escherichia coli/Shigella* strain genomes.** The upset plot depicts the pangenome composition of three bacterial strains (GN3, GN6, and GN9), utilising *E. coli k12* as a reference genome. The upper bar chart represents the number of AMR genes common across specific genome combinations. Genome combinations are indicated by the matrix below, where black dots with connecting lines denote the gene sets shared by the corresponding genomes. The left bar chart demonstrates the total quantity of identified AMR genes in each genome.

A green square with red squares

Description automatically generated

**Figure 7. Heatmap of antimicrobial resistance (AMR) genes across *Escherichia coli/Shigella* strain genomes.** The heatmap compares the presence/absence of AMR genes across three bacterial strains (GN3, GN6, and GN9), utilising *E. coli k12* as a reference genome. The colours represent gene presence (green) and absence (red).

Elongation Factor Thermo Unstable (EF-Tu) conferring resistance to pulvomycins was found to be the most common AMR gene (Table 1) and most common virulence factor. EF-Tu comprises up to 6% of the total expressed protein in *E. coli* and is one of the most common proteins in bacteria (Furano, 1975). While the primary canonical function of EF-Tu is the transportation of aminoacylated tRNAs to the ribosome (Sprinzl, 1994), diverse functions include important virulence factors across Gram-positive and Gram-negative pathogenic bacteria. EF-Tu must be retained on the extracellular surface to effect alternate virulence-associated functions, such as host extracellular matrix component adhesion (Harvey et al., 2019). They have also been targeted for therapeutic targets for antibiotics since the 1970s, however, the RGI analysis identified the genes as EF-Tu mutants conferring resistance to Pulvomycin.

**Table 1. The most frequent antimicrobial resistance (AMR) genes across *Escherichia coli/Shigella* strains**. Three strains were aligned (GN3, GN6, and GN9), utilising *E. coli k12* as a reference genome. The table contains information on the gene name, gene family, AMR mechanism, associated antibiotic drug class, and count of genes across the strains

A screenshot of a computer

Description automatically generated

*Antibiotic Efflux Pumps*

Antibiotic efflux was identified as the most common antibiotic resistance mechanism across all three strains (Figure 8). Additionally, the top three AMR gene families were the resistance-nodulation-cell division (RND), the major facilitator (MFS), and the ATP-binding cassette (ABC) antibiotic efflux pump gene families (Figure 9). Bacterial efflux pumps are generally chromosomally encoded, exhibiting conservation at both the genetic and protein levels (Blanco et al., 2016). This was followed by antibiotic target alteration, antibiotic inactivation, and reducing antibiotic permeability.

A screenshot of a computer

Description automatically generated*E. coli* utilises many multidrug resistance transporters (MDTs) that use energy derived from secondary active proton/sodium motive force for the extracellular efflux of toxic compounds from the intracellular environment (Saier, 2000). As depicted in Figure 10, the trimeric inner membrane RND transporter, AcrB (orange), is essential in energy transduction and substrate recognition (Murakami et al., 2002). Six copies of the periplasmic adaptor protein, AcrA (green), interact with the AcrB trimer, forming a sealed tubular structure and linking AcrB to TolC (red). The genes encoding this AcrAB-TolC efflux pump, *acrA* and *acrB,* form a polycistronic operon, regulated upstream of *acrA* by the *acrR* gene (Figure 10; Bay et al., 2017). Across the strains, mutations were observed in AcrR conferring resistance to ciprofloxacin, tetracycline, and ceftazidime (Table 1). RND efflux pump genes were detected 12 times across the strains, suggesting that efflux pumps are crucial in multidrug resistance by physically pumping antibiotics out of the cell, reducing intracellular concentrations (Yang et al., 2008).

**Figure 8. Antibiotic resistance mechanism distribution across *Escherichia coli/Shigella* strains.** The three strains were GN3, GN6, and GN9, utilising *E. coli k12* as a reference genome. The strains’ differences are visualised with subplots. Resistance mechanisms are visualised along the y-axis and their corresponding counts on the x-axis.

A screenshot of a graph

Description automatically generated**Figure 9. Antibiotic resistance gene family distribution across *Escherichia coli/Shigella* strains**. The three strains were GN3, GN6, and GN9, utilising *E. coli k12* as a reference genome. The strains’ differences are visualised with subplots. Gene families are visualised along the y-axis and their corresponding counts are on the x-axis.

A diagram of a structure

Description automatically generated

Figure 10. AcrAB-TolC multidrug efflux pump regulation and structure. a) Diagram presenting the potential regulatory mechanisms of *acrR* on the *acrAB* operon. b) The tripartite acrAB(Z)-TolC efflux complex crystal structure (PDB ID: 5O66). Based on research from Darby et al., (2022) and Choi et al., (2024). Created in https://BioRender.com.

*AcrR Phylogenetic Analysis*

The transcriptional regulator AcrR was identified as two paralogs, AcrR1 and AcrR2, composed of a TetR-type helix-turn-helix (HTH) domain profile and a MAATS-type transcriptional repressor, C-terminal region (Figure 12). InterPro analysis, annotated both paralogs as part of the classic multidrug efflux repressor family associated with acrEF/envCD operon regulation. These domains and families were consistent across all strains despite a lack of sequence similarity. In contrast, a maximum likelihood phylogenetic tree revealed two strongly supported clades corresponding to the AcrR1 and AcrR2 proteins (Figure 11). These clades displayed significant genetic divergence with 0.6 branch length since the last homolog. Furthermore, GN6 and GN9 form sister groups across AcrR1 (support value: 97), and AcrR2 (support value: 59). This is consistent with sub-functionalisation following gene duplication, where gene duplicates develop specialised roles whilst maintaining ancestral structural frameworks (Birchler and Yang, 2022).

A diagram of a computer

AI-generated content may be incorrect.

**Figure 11. Maximum Likelihood phylogenetic tree of AcrR paralogs across *Escherichia coli/Shigella* strains.** The three strains were GN3, GN6, and GN9, utilising *E. coli k12* as a reference genome. Phylogenetic inference was performed in MEGA11 using the T92 + I model with 500 bootstrap replicates. Evolutionary divergence is shown by branch length and bootstrap scores depict evolutionary confidence.

*AcrR Domain and Structural Analysis*

The proteins were modelled with AlphaFold2 and visualised in PyMol. All AcrR1 proteins exhibited a tightly packed α-helical architecture with minor structural differences resulting from the mutations. Structural alignment of the paralogs revealed spatial separation of 2.5 Å at the N-terminal domain (residue 1) and 32.4 Å at the C-terminal domain (residue 216). Although both homologs exhibit mutations near the C-terminal region, the extended tail on AcrR2 significantly changes the spatial orientation, suggesting regulatory divergence (Figure 12).

Sequence alignment revealed that AcrR1 contains two conserved amino acid substitutions across all three strains (T213I and N214T), localised near the C-terminal domain (Figure 13). In contrast, AcrR2 exhibited 6-7 non-synonymous mutations in the strains (Figure 14), scattered across the N-terminal DNA binding domain and clustered in the C-terminal ligand-binding region.

Further structural alignment of the AcrR2 was performed to reveal the impact of C-terminal mutations. The three models included the reference (*E. coli K12*), GN3 (6 mutations), and GN6/GN9 (7 mutations). The AcrR2 multiple missense mutations generally clustered around the C-terminal tail, however, an additional mutation at the N-terminal DNA binding domain (H27R; Figure 14) likely affects DNA recognition and operator specificity with potentially altered transcriptional repression. Notably, AcrR2 exhibits an extended C-terminal tail (15 bps) and, unlike the canonical helical architecture in AcrR1’s C-terminus, AcrR2 region appears non-helical and unstructured, forming a flexible tail.

As shown in Figure 13, at residue 214 (C-terminus), 32.4 Å was observed between GN3 and the reference, while 32.5 Å between GN6/GN9 and the reference. Additionally, there was 2.6 Å between GN3 and GN6/GN9 at this residue. Additionally, at residue 1 (N-terminus), distances of 0.6 Å between the reference and GN3, 1.2 Å between GN6/GN9 and the reference, and 1.2 Å between GN6/GN9 and GN3 were recorded. In contrast, AcrR1 presents minimal spatial change at the N-terminal region (0.6 Å) and the C-terminal region (0.7 Å). This suggests that *acrR2* is more prone to mutation and structural divergence.

This comparative analysis of the AcrR2 models highlights the dramatic rearrangement of the C-terminus, suggesting GN6/GN9 specific missense mutations, such as Q148L, may impact general conformational effects at the tail region. This raises caution of antibiotic resistance development with such conformational changes resulting from single amino acid substitutions. Moreover, the striking difference between the reference model and both GN3 and GN6/GN9 models, shown by its >30 **Å** displacement, reinforces the idea that this is a rapidly evolving protein that may undergo adaptive structural shifts in response to selective pressures.

Figure 15 shows different interaction partners between the paralogs, with acrR1 upstream of *acrA* and *acrB\_1*, whereas *acrR2* was upstream *acrE* and *acrF*. This suggests differential operon regulation and potential specialised efflux pump control. The AcrEF-TolC system is homologous to the AcrAB-TolC efflux pump, expelling many antibiotics. While *acrS* typically regulates *acrEF*, *acrR2’s* interaction with *acrE* indicates potential sub-functionalisation. Insertion sequences like IS1 or IS10 have been shown to integrate upstream of the *acrEF* operon, leading to overexpression (Olliver et al., 2005). This may impose metabolic burdens that require potential regulatory mechanisms from proteins like AcrR2. Moreover, mutations to the *acrR* gene are associated with increased *acrAB* expression and induce highly virulent infections in mouse models, leading to death (Subhadra et al., 2018).

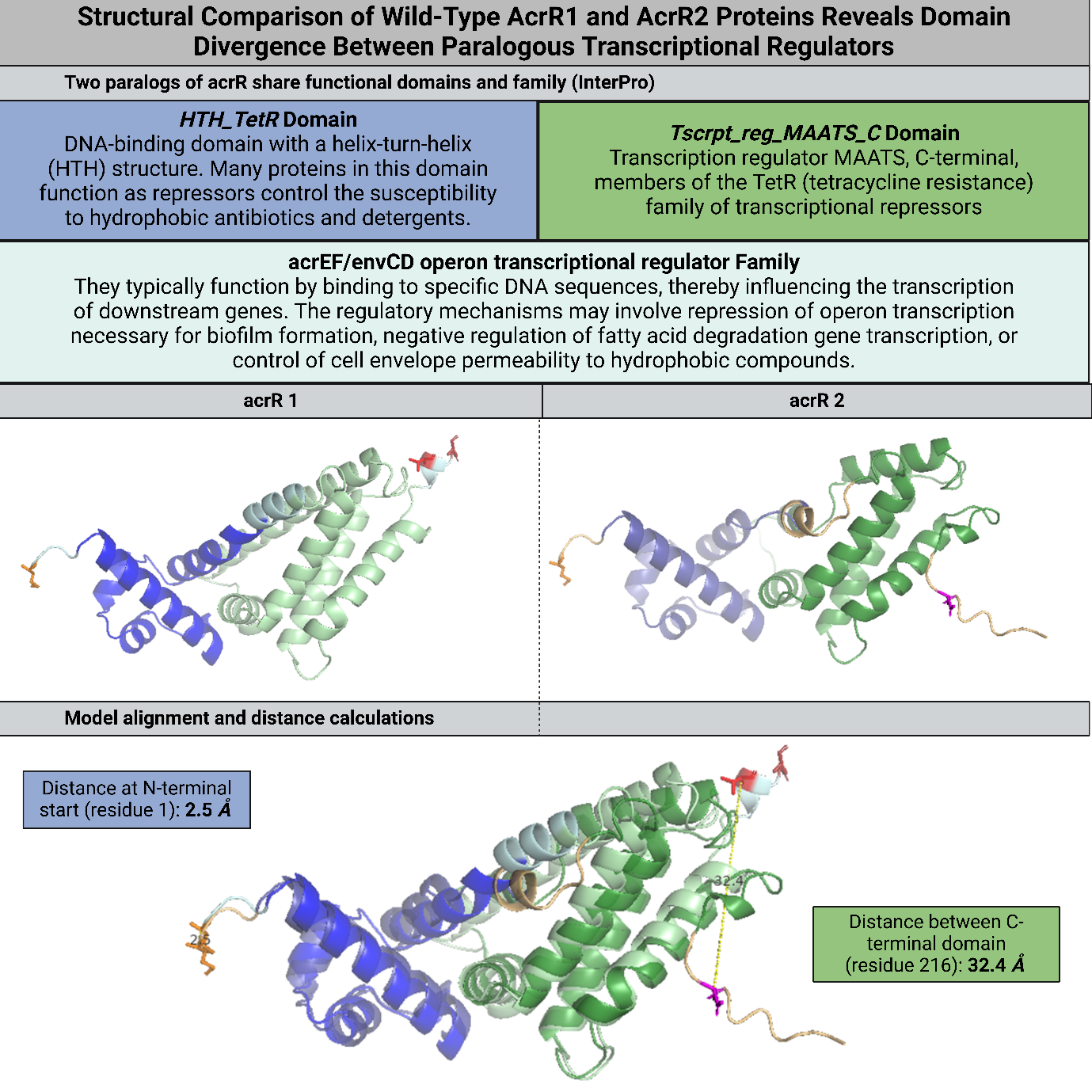


Figure 12. Structural modelling of AcrR paralogs in *Escherichia coli/Shigella* strainsreveals divergent domain architecture. InterPro was used to align AcrR1 and AcrR2 to compare the domains. Both are part of the acrEF/envCD operon transcriptional regulator family and share the HTH\_TetR\_2 DNA binding domain (blue) and the TetR\_C\_2 ligand binding domain (green). The paralogs were modelled using ColabFold v1.5.5. AcrR1 and AcrR2 were aligned and visualised with PyMOL, with mutations marked as sticks. Following structural alignment, PyMOL-based spatial measurements calculated the distances between the N-terminal and C-terminal residues. Created in https://BioRender.com.

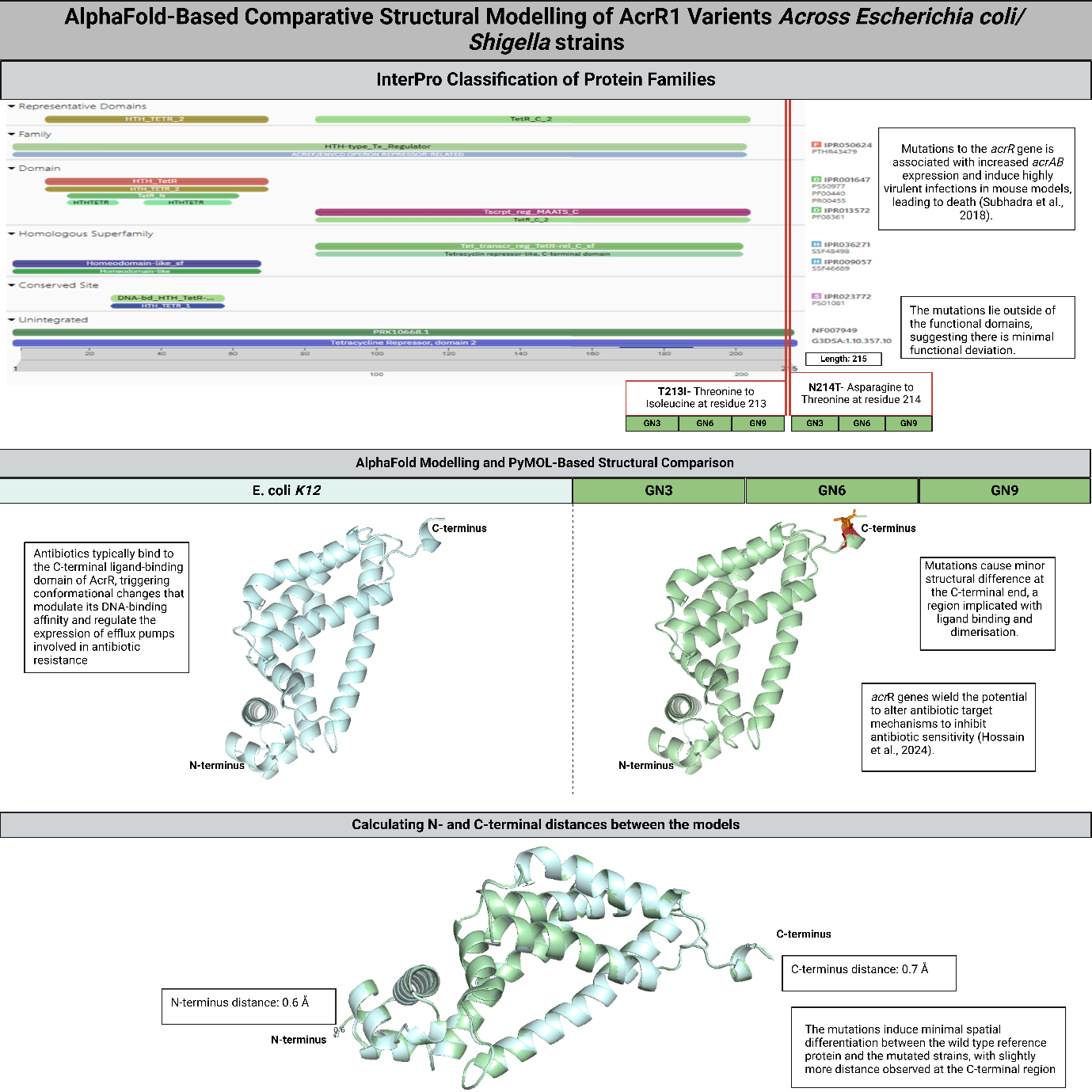


Figure 13. Structural modelling and comparative domain analysis of AcrR1 across *Escherichia coli/Shigella* strains. InterPro was used to align AcrR1 and AcrR2 to compare the domains. Both are part of the acrEF/envCD operon transcriptional regulator family and share the HTH\_TetR\_2 DNA binding domain and the TetR\_C\_2 ligand binding domain. After modelling with ColabFold v1.5.5, AcrR1 was aligned and visualised using PyMOL, highlighting mutations marked as sticks. Mutations T213I and N214T were conserved across all strains (GN3, GN6, GN9). *E. coli* K12 was used as a reference for comparison. The distance between the different model’s N- and C-terminal were calculated for spatial analysis. Created in https://BioRender.com.

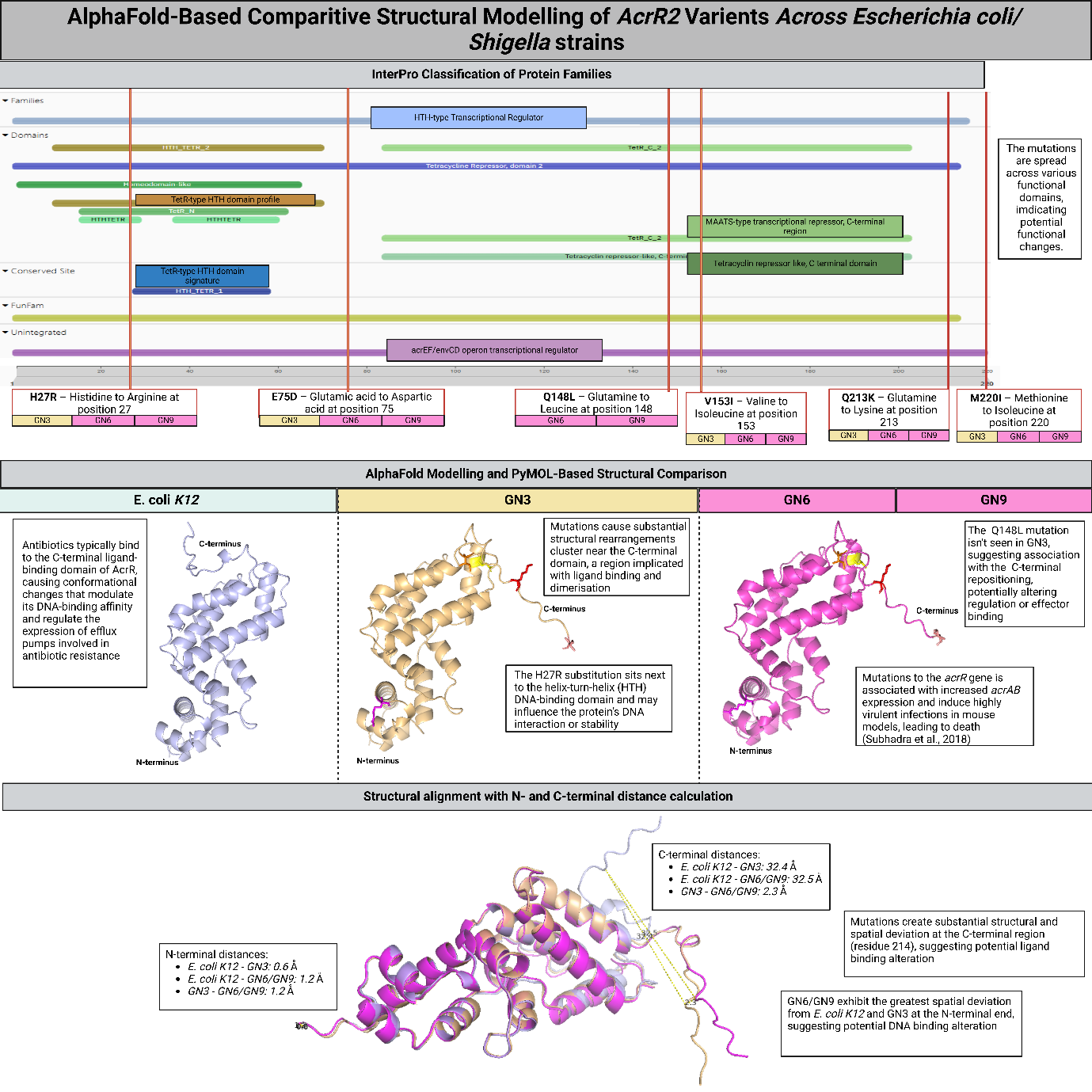


Figure 14. Structural modelling and comparative domain analysis of AcrR2 across *Escherichia coli/Shigella* strains. InterPro was used to investigate domains of all strains (*E. coli* K12, GN3, GN6, GN9). Both are part of the acrEF/envCD operon transcriptional regulator family and share the HTH\_TetR\_2 DNA binding domain and the TetR\_C\_2 ligand binding domain. After modelling with ColabFold v1.5.5, AcrR1 was aligned and visualised using PyMOL, highlighting mutations marked as sticks. Six mutations were conserved across all strains (GN3, GN6, GN9), with one mutation (Q148L) conserved only across GN6 and GN9. *E. coli* K12 was used as a reference for comparison. The distance between the different model’s N- and C-terminal were calculated for spatial analysis. Created in https://BioRender.com.

A screenshot of a computer screen

AI-generated content may be incorrect.

Figure 15. Network analysis of acrR1 and acrR2 across *Escherichia coli/Shigella* strains. Panaroo pangenome analysis was conducted across of all strains (*E. coli* K12, GN3, GN6, GN9). The core genome was visualised in Cytoscape and clustered to identify the interactions of the acrR paralogs. Created in https://BioRender.com.

*Future Perspectives*

In conclusion, this study revealed a highly conserved core genome enriched with AMR genes, abundantly associated with mutated efflux pump systems conveying resistance to numerous antibiotics. The identification of two divergent AcrR paralogs highlights the potential sub-functionalisation post-gene duplication. Furthermore, structural modelling revealed the extensive C-terminal variation and intra-domain structural rearrangements in AcrR2, suggesting a possible change in regulation and ligand binding. This may contribute to antibiotic resistance via binding site alteration, demonstrating how genetic mutations can drive AMR. A genome-wide analysis of Shigella revealed a substantial proportion of resistance genes conferring resistance through efflux pump systems, highlighting the significance of multidrug resistance. Overall, the substantial mutations in the efflux pump regulator AcrR underscore the need for sophisticated surveillance strategies and persistent structural-functional studies to anticipate resistance, mitigate resistance spread, and inform therapeutics development in *E. coli/Shigella.* Although this study utilises a limited number of replicates, the range of mutations observed highlights the need for deeper research into *Shigella* AMR mechanisms. These AMR genes are commonly embedded into mobile genetic elements, facilitating rapid spread across bacterial populations (Carattoli, 2013). Globally, we must combat the overuse and miss use of antibiotics, with future studies must focussing on *Shigella* specific research to understand the factors driving AMR growth, benefitting the development of targeted therapies to counter this global health threat.

Wordcount: 3548- 1052 (figure legends) = 2496 words

**Appendix**

See supplementary files for R analysis

**Reference list**

Agari, Y., Sakamoto, K., Kuramitsu, S. and Shinkai, A. (2012). Transcriptional Repression Mediated by a TetR Family Protein, PfmR, from Thermus thermophilus HB8. *Journal of Bacteriology*, 194(17), pp.4630–4641. doi:https://doi.org/10.1128/jb.00668-12.

Bay, D.C., Stremick, C.A., Slipski, C.J. and Turner, R.J. (2017). Secondary multidrug efflux pump mutants alter Escherichia coli biofilm growth in the presence of cationic antimicrobial compounds. *Research in Microbiology*, 168(3), pp.208–221. doi:https://doi.org/10.1016/j.resmic.2016.11.003.

Birchler, J.A. and Yang, H. (2022). The multiple fates of gene duplications: deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *The Plant Cell*, 34(7). doi:https://doi.org/10.1093/plcell/koac076.

Blanco, P., Hernando-Amado, S., Reales-Calderon, J., Corona, F., Lira, F., Alcalde-Rico, M., Bernardini, A., Sanchez, M. and Martinez, J. (2016). Bacterial Multidrug Efflux Pumps: Much More Than Antibiotic Resistance Determinants. *Microorganisms*, 4(1), p.14. doi:https://doi.org/10.3390/microorganisms4010014.

Carattoli, A. (2013). Plasmids and the spread of resistance. *International Journal of Medical Microbiology*, 303(6-7), pp.298–304. doi:https://doi.org/10.1016/j.ijmm.2013.02.001.

Choi, N., Choi, E., Cho, Y.-J., Kim, M.J., Choi, H.W. and Lee, E.-J. (2024). A shared mechanism of multidrug resistance in laboratory-evolved uropathogenic *Escherichia coli*. *Virulence*, 15(1). doi:https://doi.org/10.1080/21505594.2024.2367648.

Cuthbertson, L. and Nodwell, J.R. (2013). The TetR Family of Regulators. *Microbiology and Molecular Biology Reviews*, [online] 77(3), pp.440–475. doi:https://doi.org/10.1128/mmbr.00018-13.

Darby, E.M., Trampari, E., Siasat, P., Gaya, M.S., Alav, I., Webber, M.A. and Blair, J.M.A. (2022). Molecular Mechanisms of Antibiotic Resistance Revisited. *Nature Reviews Microbiology*, [online] 21(5), pp.280–295. doi:https://doi.org/10.1038/s41579-022-00820-y.

Dyson, H.J. and Wright, P.E. (2016). Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Coactivators CREB-binding Protein (CBP) and p300. *Journal of Biological Chemistry*, 291(13), pp.6714–6722. doi:https://doi.org/10.1074/jbc.r115.692020.

Furano, A.V. (1975). Content of elongation factor Tu in Escherichia coli. *Proceedings of the National Academy of Sciences*, 72(12), pp.4780–4784. doi:https://doi.org/10.1073/pnas.72.12.4780.

Gordienko, E.N., Kazanov, M.D. and Gelfand, M.S. (2013). Evolution of Pan-Genomes of Escherichia coli, Shigella spp., and Salmonella enterica. *Journal of Bacteriology*, [online] 195(12), pp.2786–2792. doi:https://doi.org/10.1128/jb.02285-12.

Harvey, K.L., Jarocki, V.M., Charles, I.G. and Djordjevic, S.P. (2019). The Diverse Functional Roles of Elongation Factor Tu (EF-Tu) in Microbial Pathogenesis. *Frontiers in Microbiology*, [online] 10. doi:https://doi.org/10.3389/fmicb.2019.02351.

Hossain, Md.A., Al Amin, Md., Khan, Md.A., Refat, Md.R.R., Sohel, M., Rahman, M.H., Islam, A. and Hoque, M.N. (2024). Genome-Wide Investigation Reveals Potential Therapeutic Targets in Shigella spp. *BioMed Research International*, 2024, pp.1–14. doi:https://doi.org/10.1155/2024/5554208.

Lan, R. and Reeves, P.R. (2002). Escherichia coli in disguise: molecular origins of Shigella. *Microbes and Infection*, 4(11), pp.1125–1132. doi:https://doi.org/10.1016/s1286-4579(02)01637-4.

Liu, B., Zheng, D., Jin, Q., Chen, L. and Yang, J. (2018). VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Research*, 47(D1), pp.D687–D692. doi:https://doi.org/10.1093/nar/gky1080.

Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L. and Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *PLOS Computational Biology*, 14(1), p.e1005944. doi:https://doi.org/10.1371/journal.pcbi.1005944.

Murakami, S., Nakashima, R., Yamashita, E. and Yamaguchi, A. (2002). Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature*, 419(6907), pp.587–593. doi:https://doi.org/10.1038/nature01050.

Olliver, A., ValléM., Chaslus-Dancla, E. and Cloeckaert, A. (2005). Overexpression of the Multidrug Efflux Operon acrEF by Insertional Activation with IS1 or IS10 Elements in Salmonella enterica Serovar Typhimurium DT204 acrB Mutants Selected with Fluoroquinolones. *Antimicrobial Agents and Chemotherapy*, 49(1), pp.289–301. doi:https://doi.org/10.1128/aac.49.1.289-301.2005.

Parks, D.H., Chuvochina, M., Reeves, P.R., Beatson, S.A. and Hugenholtz, P. (2021). Reclassification of *Shigella* species as later heterotypic synonyms of *Escherichia coli* in the Genome Taxonomy Database. doi:https://doi.org/10.1101/2021.09.22.461432.

Poirel, L., Madec, J.-Y., Lupo, A., Schink, A.-K., Kieffer, N., Nordmann, P. and Schwarz, S. (2018). Antimicrobial Resistance in Escherichia coli. *Microbiology spectrum*, [online] 6(4). doi:https://doi.org/10.1128/microbiolspec.ARBA-0026-2017.

Saier, M.H. (2000). A Functional-Phylogenetic Classification System for Transmembrane Solute Transporters. *Microbiology and Molecular Biology Reviews*, 64(2), pp.354–411. doi:https://doi.org/10.1128/mmbr.64.2.354-411.2000.

Sharma, R. and Sanjay Kumar Dubey (2019). Computational Management of Alignment of Multiple Protein Sequences Using ClustalW. *Advances in intelligent systems and computing*, pp.177–186. doi:https://doi.org/10.1007/978-981-15-0029-9\_14.

Sprinzl, M. (1994). Elongation factor Tu: a regulatory GTPase with an integrated effector. *Trends in Biochemical Sciences*, [online] 19(6), pp.245–250. doi:https://doi.org/10.1016/0968-0004(94)90149-x.

Subhadra, B., Kim, J., Kim, D.H., Woo, K., Oh, M.H. and Choi, C.H. (2018). Local Repressor AcrR Regulates AcrAB Efflux Pump Required for Biofilm Formation and Virulence in Acinetobacter nosocomialis. *Frontiers in Cellular and Infection Microbiology*, 8. doi:https://doi.org/10.3389/fcimb.2018.00270.

Sweta Padma Routray, Sahoo, S., Kumar, S., Shah, S. and Tripti Swarnkar (2024). Web-Based Tools Validation for Antimicrobial Resistance Prediction: An Empirical Comparative Analysis. *SN Computer Science/SN computer science*, 5(1). doi:https://doi.org/10.1007/s42979-023-02460-2.

Tompa, P. (2002). Intrinsically unstructured proteins. *Trends in Biochemical Sciences*, 27(10), pp.527–533. doi:https://doi.org/10.1016/s0968-0004(02)02169-2.

Tompa, P. (2012). Intrinsically disordered proteins: a 10-year recap. *Trends in Biochemical Sciences*, [online] 37(12), pp.509–516. doi:https://doi.org/10.1016/j.tibs.2012.08.004.

Werten, S., Waack, P., Palm, G.J., Virolle, M. and Hinrichs, W. (2022). Crystal structures of free and ligand-bound forms of the TetR/AcrR-like regulator SCO3201 from Streptomyces coelicolor suggest a novel allosteric mechanism. *The FEBS Journal*. doi:https://doi.org/10.1111/febs.16606.

Yang, H., Duan, G., Zhu, J., Lv, R., Xi, Y., Zhang, W., Fan, Q. and Zhang, M. (2008). The AcrAB-TolC Pump Is Involved in Multidrug Resistance in Clinical*Shigella flexneri*Isolates. *Microbial Drug Resistance*, 14(4), pp.245–249. doi:https://doi.org/10.1089/mdr.2008.0847.

Yang, Z., Zeng, X., Zhao, Y. and Chen, R. (2023). AlphaFold2 and its applications in the fields of biology and medicine. *Signal Transduction and Targeted Therapy*, [online] 8(1). doi:https://doi.org/10.1038/s41392-023-01381-z.