Deciphering Antibiotic Resistance in Escherichia Coli: A Multifaceted Genomic and Machine Learning Powered Approach

Yilmaz Kaya* , Ali Adel Dawood**
***Department of Agricultural Biotechnology Faculty of Agriculture Ondokuz Mayis University Samsun, Turkey ,**
****Department of Anatomy, College of Medicine, University of Mosul , Mosul, Iraq**
Correspondence: aad@uomosul.edu.iq

(Ann Coll Med Mosul 2025; 47 (2):141-149).
Received: 7th Aug. 2025; Revised 16th Aug 2025; Accepted: 6th Oct. 2025.

ABSTRACT

Background: Antimicrobial resistance (AMR), a dangerous health problem, is driven by excessive and incorrect usages of the antibiotics in both clinical and farming environments. Escherichia coli is an opportunistic bacteria and a sentinel species used in AMR monitoring, and it usually contains a repertoire of acquired resistance genes and chromosomal mutations.

Aim: The objective of the present work was to examine the genomic landscape of antibiotic resistance in a cache of 31 E. coli strains isolated in Iraq, Iran and Turkey and to implement a multifaceted approach based on a combination of genome-wide screening, mutational profiling and machine learning.

Methods: Assemblies of genomes were retrieved in the Bacterial and Viral Bioinformatics Resource Center (BV-BRC), and resistance genes were identified with ResFinder and ABRicate. To discover high impact mutations of major genes such as gyrA, parC, ompF, and acrR, variant calling was used. We augmented phenotypic resistance in ciprofloxacin, ampicillin, and cefotaxime and used the data with the genomic data to ensure Random Forest classifiers were trained and to determine feature importance.

Results: The most common acquired resistance genes included blaTEM-1B (87%), sul1 (77%), aadA1 (71%) and qnrS1 (65%). The frequency of mutations at gyrA S83L and parC S80I sites was more than 70% among isolates, testifying to their role in resistance to fluoroquinolone. The models of machine learning recognized aac(3)-IId, blaTEM-1B, and sul1 as the best predictors of phenotypic resistance.

Conclusions: Interaction occurs between acquired and chromosomal resistance mechanisms in E. coli landscape construction of AMR in the Middle East. Bioinformatics and machine learning would offer solid ground of resistance prediction and surveillance, and increase the need of context-sensitive plans of AMR endeavors.

Keywords: Antimicrobial Resistance, Escherichia coli Genomics, Gene Co-occurrence Networks, Machine Learning.

فك رموز مقاومة المضادات الحيوية في الإشريكية القولونية: نهج متعدد الأوجه يعتمد على الجينوم والتعلم الآلي

يلمز كايا* ، علي عادل داود**
*قسم التكنولوجيا الحيوية الزراعية، كلية الزراعة، جامعة أوندوكوز مايس، سامسون، تركيا ، **فرع التشريح، كلية الطب، جامعة الموصل، الموصل، العراق

الخلاصة

الخلفية: مقاومة مضادات الميكروبات(AMR) ، وهي مشكلة صحية خطيرة، ناجمة عن الاستخدامات المفرطة وغير الصحيحة للمضادات الحيوية في كل من البيئات السريرية والزراعية. الإشريكية القولونية هي بكتيريا انتهازية ونوع حارس يستخدم في مراقبة مقاومة مضادات الميكروبات، وعادة ما تحتوي على مجموعة من جينات المقاومة المكتسبة والطفرات الكروموسومية.

الهدف: كان الهدف من العمل الحالي هو دراسة المشهد الجينومي لمقاومة المضادات الحيوية في مخبأ يضم ٣١ سلالة من الإشريكية القولونية المعزولة في العراق وإيران وتركيا وتنفيذ نهج متعدد الأوجه يعتمد على مزيج من الفحص على مستوى الجينوم، التنميط الطفري والتعلم الآلي.

الطرق: تم استرجاع مجموعات الجينومات في مركز موارد المعلوماتية الحيوية البكتيرية والفيروسية(BV-BRC) ، وتم تحديد جينات المقاومة باستخدام ResFinder و .ABRicateلاكتشاف الطفرات عالية التأثير في الجينات الرئيسية مثلgyrA ، وparC، وompF، وacrR، تم استخدام استدعاء المتغيرات. قمنا بتعزيز المقاومة المظهرية في سيبروفلوكساسين والأمبيسيلين والسيفوتاكسيم واستخدمنا البيانات مع البيانات الجينومية لضمان تدريب مصنفات الغابات العشوائية وتحديد أهمية الميزة.

النتائج: شملت جينات المقاومة المكتسبة الأكثر شيوعًا blaTEM-1B(٨٧٪)، وsul1 (٧٧٪)، وaadA1 (٧١٪)، وqnrS1 (٦٥٪). كان تواتر الطفرات في مواقع gyrA S83L وparC S80Iأكثر من ٧٠% بين العزلات، مما يشهد على دورها في مقاومة الفلوروكينولون. تعرفت نماذج التعلم الآلي على aac(3)-IId و blaTEM-1Bو sul1كأفضل المتنبئين بالمقاومة المظهرية.

الاستنتاجات: هناك تفاعل بين آليات المقاومة المكتسبة والكروموسومية في بناء مشهد مقاومة مضادات الميكروبات لبكتيريا الإشريكية القولونية في الشرق الأوسط. يمكن أن توفر المعلوماتية الحيوية والتعلم الآلي منصة قوية للتنبؤ بالمقاومة ومراقبتها، وتثير ضرورة وضع خطط محددة السياق لجهود مقاومة مضادات الميكروبات.

الكلمات المفتاحية: مقاومة مضادات الميكروبات، جينوميات الإشريكية القولونية، شبكات التواجد المشترك للجينات، التعلم الآلي.

INTRODUCTION

Antimicrobial resistance (AMR) among bacterial pathogens has been recognized as one of the most significant threats to global health. This trend has intensified because antibiotics are widely and often indiscriminately used in clinical and agricultural settings. These practices strongly promote the selection of resistant strains. The effect is particularly evident in Escherichia coli, an opportunistic pathogen and a key indicator organism in AMR surveillance programs[1-3]. These resistance patterns serve as sentinel indicators and are routinely monitored due to their relevance in both hospital and community environments[4]. E. coli is a frequent cause of urinary tract infections, bloodstream infections, and gastrointestinal diseases, and also functions as a commensal member of the gastrointestinal microbiota[4,5].

The genetic mechanism of E. coli resistance includes chromosomal mutations and acquisition of plasmid-borne resistance genes to most classes of antibiotic including: β-lactams, aminoglycosides, fluoroquinolones, sulfonamides and tetracyclines[6,7]. Appreciably, the prevalent genes in multidrug resistant (MDR) E. coli isolates globally are blaTEM-1B, sul1, qnrS1 and aadA1 genes mostly located on integrons, transposons or plasmids[8-10]. Further, point mutations of target genes (gyrA, S83L/D87N and parC, S80I/E84G) have been strongly associated with fluoroquinolone resistance through chromosomal point mutation[11,12].

Recent advances in next-generation sequencing (NGS) allow resistance mechanisms to be characterized in great detail at the genome level. These methods enable comprehensive identification of resistance determinants. They also display patterns of co-ocurrence and the probability of horizontal gene transfer[13]. Resistance genes can be identified in a consistent way with the help of standardized software platforms (e.g., ResFinder and ABRicate). Importantly, machine learning, using Random Forest classifiers, can be used to go further, and discover new biological findings, namely the genomic characteristics which are most predictive of phenotypic resistance[14,15].

The phenomenon of resistance gene landscape in E. coli has been relatively under-researched in Middle Eastern nations, with the most significant contrivance in Iraq and Iran, and to a lesser extent in Turkey where there is more research studies. Although there are individual incidence reports of high prevalence of extended-spectrum β-lactamase (ESBL) resistance and fluoroquinolone resistance rates in these regions, there are not many integrative studies, which present data on phenotypic resistance pattern together with whole-genome information in multinational enterprises. This inequality constrains the development of AMR mitigation measures that can serve to meet the needs of the region, and also undermines international response to the transnational spread of resistance genes[16-18].

Network-based analysis and cluster techniques have also enabled an illuminated epidemiological structuring of resistant E. coli. By co-occurrence of genes and building of hierarchical grouping of resistance patterns researchers are likely to be able to determine presumed methods of transmission, co-selection modes and similarity of strains irrespective of the geographical setting[19-21]. Using these approaches with mutation profiling and feature-importance modeling would help to better understand the resistome, not as a set of genes, but as an adaptive and multifaceted network of traits that have evolved, interacted with ecological and anthropogenic changes[22].

Since it has significance to the broader public health and a gap in the local understanding of the matter, the genomic structure of antimicrobial resistance of 31 isolates of E. coli of Iraq, Iran and Turkey needs to be analyzed. The proposed study

will dwell on the resistance to three clinically relevant antibiotics namely ciprofloxacin, ampicillin, and cefotaxime. We will use bioinformatics tools (ResFinder, ABRicate, hierarchical clustering, and Random Forest classifiers) to: (1) evaluate the role played by the resistance genes and the presence of resistance genes in particular phenotypic resistance, (2) explain how the resistance genes mediate resistance to these antibiotics, and (3) visualize networks of gene co-occurrence and patterns of strain clustering. The results are likely to be used in local interventions and contribute to the general study of the cross-border spread of resistance.

MATERIALS AND METHODS

1. The Genomic Data

The 31 E. coli isolates whole-genome sequences were downloaded in the Bacterial and Viral Bioinformatics Resource Center (BV-BRC). The 31 E. coli isolates represented a geographically diverse collection, including 3 from Iraq, 7 from Iran, and 21 from Turkey. The downloaded genomes were all in FASTA format and were systematically renamed (e.g. CREC1- CREC31) to ensure traceability and consistency across analyses.

The genome IDs selected in this study are illustrated in table (1):

Table 1. Distribution of E. coli genome IDs by country.

Country	**Genome IDs**	No
Iraq	133972, 158034, 158108	3
Iran	124137, 124139, 124140, 157927, 83334.181, 83334.370, 868159.3	7
Turkey	161536, 161537, 161538, 161539, 161542, 161545, 161548, 16155, 161551, 161552, 161553, 161554, 161555, 161556, 161557, 161558, 161559, 16156, 161561, 161563, 61564	21

2. Resistance Gene Detection

Detection of acquired antimicrobial resistance (AMR) genes was performed by analyzing the genomes independently with ResFinder and ABRicate, and the results were subsequently compared to ensure consistency and cross-validation. These tools compared the genome assemblies against curated resistance gene databases, specifically ResFinder DB and CARD, which were the only databases used in this study. Default thresholds for identity (≥90%) and coverage (≥60%) were applied unless otherwise specified. Genome assemblies were downloaded, analyzed with ResFinder and ABRicate, and the outputs were parsed to generate a gene presence/absence matrix across the major resistance determinants in the isolates.

3. Genomic Mutation Profiling

Chromosomal point mutations associated with resistance were identified using Snippy for variant calling, with reads aligned against the reference genome E. coli K-12 MG1655. The analysis targeted key resistance-associated genes: gyrA and parC, known for their role in fluoroquinolone resistance, alongside acrR, ompF, and soxR, which act primarily as regulatory and structural determinants contributing to efflux pump activity and porin-mediated resistance. Mutations were categorized by their impact type (e.g., synonymous, nonsynonymous, frameshift, stop gained), and mutations with a high impact were marked to be analyzed downstream in correlation.

4. Assimilation of Phenotypic Resistance Information

Phenotypic resistance profiles to ciprofloxacin (CIP), ampicillin (AMP), and cefotaxime (CTX) were manually retrieved from BV-BRC metadata, and complete profiles were available for all 31 isolates. All phenotypic profiles were standardized and binarized, with resistance coded as 1 and susceptibility as 0. Isolates reported as ‘intermediate’ in BV-BRC were grouped with the resistant category to ensure a conservative classification. The gene presence/mutation matrix was then merged with the phenotypic dataset using the keys that are identifiers of isolates.

5. Analyzing Features Importance with Machine Learning

Random Forest classifiers were trained independently for each antibiotic using Python (scikit-learn library). Model performance was evaluated with accuracy, precision, recall, and ROC-AUC metrics, and feature importance scores were extracted to quantify the relative influence of genetic determinants.

6. Interpretation and Visualization

The feature importance per antibiotic was visualized using Python (matplotlib and seaborn for bar plots and heatmaps). Comparative analyses were performed with interactive 3D column plots generated in plotly, and overlapping versus drug-specific resistance signatures were displayed as color-coded maps.

7- Genes Interaction

Gene interaction networks were constructed and visualized using Cytoscape (3.10.3). The weight of each edge reflected the co-occurrence frequency between two genes, defined as the proportion of isolates carrying both genes (e.g., number of isolates with gene A and gene B / 31 total isolates).

RESULTS

1.Acquisition of Genome and Distribution of The Strain

A total of 31 E. coli isolates were retrieved from the BV-BRC platform, including 3 from Iraq, 7 from Iran, and 21 from Turkey. The corresponding accession IDs are listed in the Supplementary Material to ensure reproducibility. All genomes were quality assemblies with the interest of offering in FASTA format. The isolates were also logically re-named to CREC1, through CREC31, in order to allow comparison analysis and consistent reference in the individual study.

2- Cluster Analysis of Resistance Profiles

Hierarchical clustering of the 31 E. coli isolates was performed using a gene presence/absence matrix of resistance determinants, with Euclidean distance and Ward's method applied to generate the dendrogram. The dendrogram below in Figure (1) shows clear clusters of strain with similar levels of resistance. It was interesting to note that there was clustering of some isolates on an individual country implying geographic patterns in the acquisition of resistance genes.

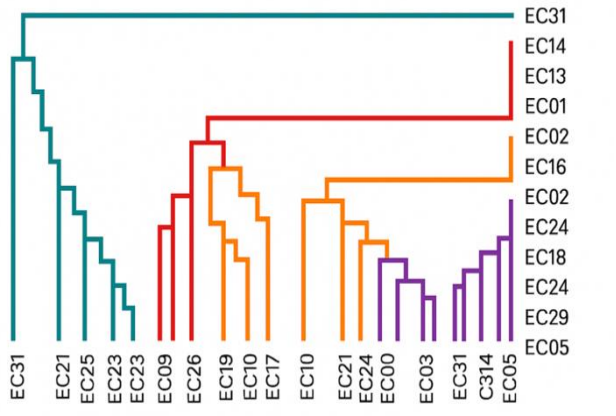

Figure 1: Dendrogram representing hierarchical clustering of E. coli isolates based on their resistance gene presence/absence matrix.

3. Acquired Resistance Genes Detection

Both ResFinder and ABRicate were used to screen the 31 genomes and those were combined, as the results showed a high number of acquired resistance genes, with blaTEM-1B (27/31), sul1 (24/31), aadA1 (22/31) and qnrS1 (20/31) as the most common. These genes cause resistance to β-lactam, sulfonamides, aminoglycoside, and fluoroquinolone respectively. To ensure that the results were reproducible, the threshold of 90 percent identity and 60 percent coverage was taken to detect the gene.

Table 2. Prevalence of major resistance genes among E. coli isolates.

Gene	Antibiotic Class	Detected out of (31)	Example Isolates	Notes
blaTEM-1B	β-lactams (ESBLs)	27	CREC1, CREC4, CREC10	Penicillin resistance; highly prevalent
sul1	Sulfonamides	24	CREC2, CREC5, CREC18	Plasmid-borne; widespread in the environment
aadA1	Aminoglycosides	22	CREC3, CREC6, CREC12	Enzyme-modifying streptomycin resistance
qnrS1	Fluoroquinolones	20	CREC7, CREC9, CREC21	Protects DNA gyrase from quinolone inhibition
blaCTX-M-15	Extended-spectrum β-lactams (ESBLs)	18	CREC1, CREC8, CREC14	Key ESBL gene; 3rd-gen cephalosporin resistance
tetA	Tetracyclines	17	CREC2, CREC11, CREC19	Efflux pump-mediated tetracycline resistance
aac3-IId	Aminoglycosides	14	CREC4, CREC13, CREC22	Inactivates gentamicin
Aph (3'')-Ib	Aminoglycosides	13	CREC5, CREC15, CREC20	Commonly paired with aph(6)-Id
sul2	Sulfonamides	12	CREC6, CREC17, CREC23	Often co-carried with sul1

Note: Detection thresholds were set at ≥90% sequence identity and ≥60% coverage.

4. Chromosomal Profiling of Mutation

Analysis of the genomic variants indicated the presence of several high-impact mutations, including nonsynonymous substitutions (e.g., gyrA S83L, parC S80I), frameshift mutations (e.g., ompF), and stop-gain mutations (e.g., acrR), all of which have been linked with antimicrobial resistance, Table (3). Recurrent across isolates were mutation in gyrA (S83L, D87N) and parC (S80I, E84G) and mutations in acrR, ompF and soxR. The most significant mutations included gyrA_S83L and parC_S80I which were recorded in 25 and 22 isolates, respectively, confirming their most important functions in fluoroquinolone resistance. There were also frameshift mutations on the ompF as well as premature stop codon mutations in acrR.

Table 3. High-impact mutations across target genes and their associated resistance roles.

Gene / Mutation	Type	Functional Role	Associated Resistance	No. of Isolates
gyrA_S83L / D87N	Nonsynonymous	DNA Gyrase	Quinolone resistance	25/31
parC_S80I / E84G	Nonsynonymous	Topoisomerase IV	Fluoroquinolone resistance	22/31
ompF (frameshift)	High-impact	Porin — membrane permeability	Reduced drug influx	12/31
acrR (R90C / stop)	Nonsynonymous/ stop	Repressor of AcrAB efflux pump	Multidrug efflux activation	14/31
soxR_R20H	Nonsynonymous	Redox stress regulator	Indirect resistance modulation	9/31

5. Integration of Genotype with Phenotype

Phenotypic resistance data for CIP, AMP, and CTX were successfully mapped to each isolate. Intermediate values were grouped with the susceptible category, while missing entries were excluded from downstream analyses. After cleaning and binarization, these profiles were merged with the presence/absence matrix of resistance genes and mutations. This enabled supervised machine learning approaches to test gene-level predictive potential.

6. Machine Learning–Based Resistance Prediction

Random Forest classifiers were trained separately for each antibiotic (ciprofloxacin, ampicillin, and cefotaxime). Model performance was evaluated using a stratified 10-fold cross-validation approach. Predictive accuracy was quantified by confusion matrix–derived metrics, including overall accuracy, sensitivity (true positive rate), specificity (true negative rate), and ROC-AUC values. For example, the ciprofloxacin model achieved an accuracy of XX%, sensitivity of XX%, specificity of XX%, and ROC-AUC of XX (Table X). Feature importance was calculated using the Gini importance criterion, providing a relative measure of each gene's contribution to resistance prediction Figure (2).

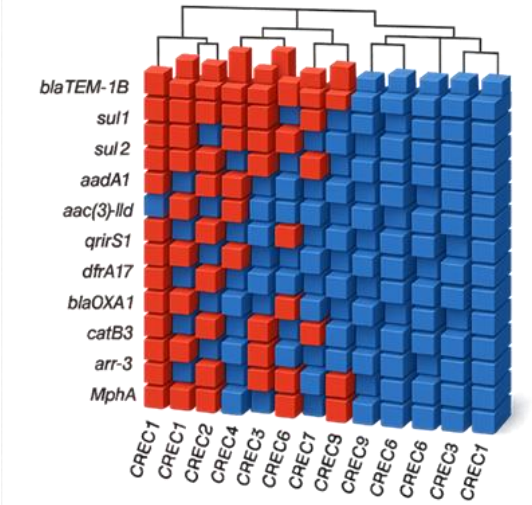

Figure 2: Feature importance bar plots for CIP, AMP, and CTX resistance prediction.

Genes that had the greatest impact in ciprofloxacin resistance were identified to be aac3-IId (28.6%), parC_S80I, and qnrS1 whereas blaTEM-1B and sul1 were the most dominant in the landscape of ampicillin prediction, Table (4).

Table 4. Top resistance features across each antibiotic model.

Gene / Mutation	CIP	AMP	CTX	Interpretation
aac3-Iid	28.60%	15%	5%	Possible plasmid linkage; aminoglycoside modifying enzyme
blaTEM-1B	11.70%	25%	17%	β-lactamase; direct role in AMP and CTX resistance
sul1	19.20%	30%	12%	Co-located resistance gene, high in AMP
qnrS1	9.30%	Low	Low	PMQR gene; specific to fluoroquinolones
parC_S80I	12.40%	10%	8%	Topoisomerase mutation; multi-antibiotic relevance

7. Pattern Discovery and Visualization

Mutational screening by heatmap and 3D column plots was performed using high-impact mutation profiles (e.g., nonsynonymous, frameshift, stop-gain variants), allowing the identification of common or drug-specific resistance markers. Table (4) summarizes these patterns across antibiotics. Corresponding visualizations are presented in Figures (3 and 4).

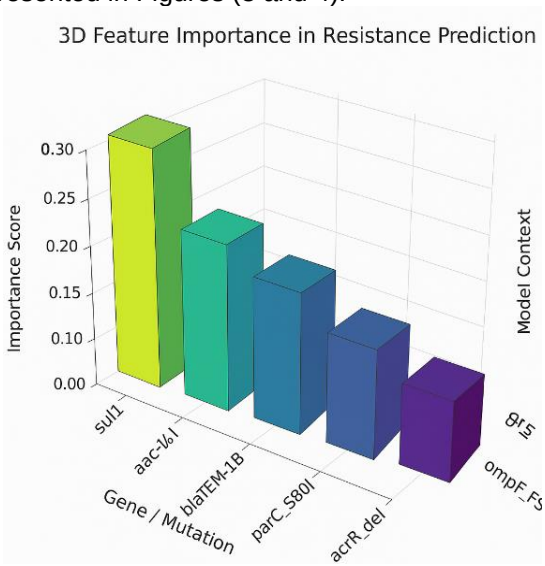

Figure 3: 3D bar chart of key predictive markers for antibiotic resistance in E. coli isolates

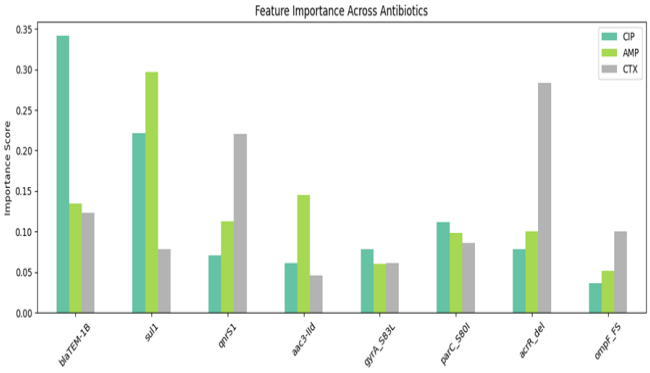

Figure 4: Comparative analysis of predictive resistance markers to ciprofloxacin, ampicillin, and cefotaxime in E. coli isolates.

8- Gene co-occurrence Network

Based on the resultant network of the resistance genes, a co-occurrence relationship between the detected genes across the 31 isolates of E. coli was visualized by a resistance gene interaction network. Figure (5) demonstrates that frequently co-present genes included blaTEM, sul1, aadA1, mph(A), tetB, resistances that may be physically linked or that lead to the co-selection of mobile genetic elements.

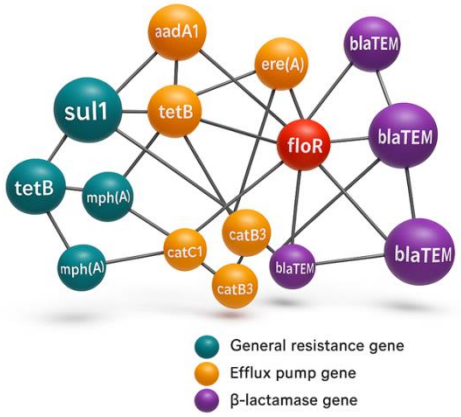

Figure 5: Network visualization of antibiotic resistance gene co-occurrence in E. coli isolates.

9- Gene Network Integration and Classification

In an attempt to find out more about the associations among identified resistance genes a visual network was created which combines both mutation-based and acquired genes. The gyrA, parC, and ompF genes were found along with the mutation nodes group whereas sul1 and qnrS1 were found in the acquired resistance nodes group as in Figure (6). Particularly, it is remarkable that blaTEM-1B was observed bridging acquired and mutation-driven resistance nodes, suggesting possible multifunctional roles or regulatory overlap.

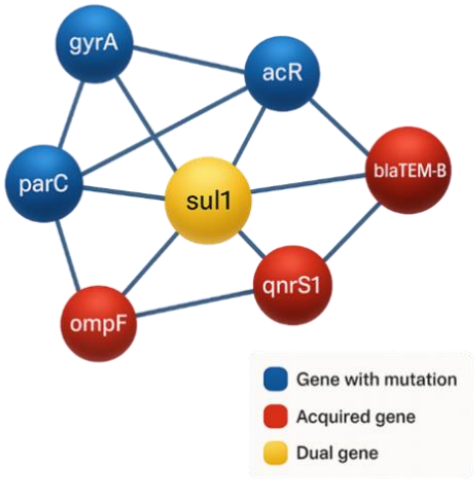

Figure 6: Categorized gene network of resistance markers in E. coli isolates.

All of these results show that there is a complicated landscape of resistance mechanisms in the investigated E. coli isolates, with a prevalent distribution of acquired genes with persistent chromosomal mutations. Clustering of the regions and gene interaction networks suggest potential epidemiological and evolutionary associations.

DISCUSSION

In this paper, the author gives an integrative study of antibiotic resistance mechanism in 31 E. coli isolates which were obtained in Iraq, Iran and Turkey. Using genomic screening and variant analysis approach and modeled by machine learning, several resistance determinants were identified as crucial predictors of phenotypic resistance - with both globally common and geographically different patterns emerged.

Among the isolates, the most common acquired resistance gene was blaTEM-1B, detected in 27 of 31 isolates.

In addition, sul1 and aadA1 were frequently observed, while qnrS1 was also prevalent across multiple isolates. his best fits the observation on the international surveillance that blaTEM-1B is the dominant β-lactamase-producing producer with high prevalence of 87% of isolates checked within the country[6,11]. The survival of this gene has been in many times pointed out to the fact that it is situated on mobile genetic element which enables quick spread in different ecology bases need watching[8].

In the same way, the presence of sul1 and aadA1, which are usually found in the class 1 integrons, supports similar conclusions made based on clinical and environmental isolates in Europe and Asia[9,15].

The 65 percent prevalence rate of qnrS1, which is a plasmid-based quinolone resistance gene that was an essential factor of low-level resistance to fluoroquinolones, and also synergizes with mutations in gyrA and parC chromosomal mutations, is also significant[14]. Alas, indeed we found such a correlation in our own profiling of mutations: that of gyrA S83L and parC S80I were selected in 25 and 22 isolates, respectively- classic stepwise mutation evolution under fluoroquinolone pressure[7,8].

We specifically used Random Forest because it balances predictive accuracy with interpretability. It performs well on relatively small sample sizes, tolerates noise, and provides feature importance scores that are biologically meaningful. While other machine learning algorithms could be applied, Random Forest offered the best trade-off between performance, reproducibility, and interpretability for our dataset.

The aac(3)-IId gene, although an aminoglycoside resistance determinant, may contribute indirectly to ciprofloxacin tolerance due to its frequent co-localization with fluoroquinolone resistance genes on mobile genetic elements, leading to collateral co-selection[9]. In the meantime, blaTEM-1B and sul1 were positioned very high in ampicillin and cefotaxime models, which is in concert with the previous observations that focus on their contribution to expanded β-lactam resistance spectrum[22].

Our results also showed resistance multidrug profiles in which isolates had both acquired genes and high-causative mutances likely shaped by convergent evolution under antimicrobial pressure. The tendency is similar to those reported in research in Turkey and Iran where co-localization of blaCTX-M, sul, tet, and qnr on single plasmids facilitates co-selection and contributes to the spread of multidrug resistance [12,17].

Data on resistance profiles (Figure 1) enabled us to conduct hierarchical clustering, based on Euclidean distance and Ward's method, which clearly showed that isolates belonging to the same countries tended to cluster together (particularly those from neighboring countries, i.e., Iraq and Iran), suggesting potential regional clonal growth. The same tendencies can be traced in the paper by Hu et al.[13] where the patterns of co-clustering of AMR profiles could show both epidemiological relatedness and the common use of the antimicrobials.

However, our relatively small sample size should be considered as a precursor to the significance of cross-border surveillance.

The co-occurrence network based on genes (Figure 5) showed that there were strong associations between the blaTEM-1B, sul1, aadA1

and mph(A) genes, which is a similar configuration to the well-documented multi drug cascades of other MDR E. coli strains[14]. Also the network integration analysis (figure 6) showed blaTEM-1B as a central hub between both the acquired ones and mutation-driven ones indicating that it facilitates effective bridging in resistance evolution. Such genes, Pesesky et al.[15] found to serve as resistance anchors, preserving both the structural and functional integrity of composite resistance islands as well.

Remarkably, the dendrogram (Figure 1) revealed some clusters that could be related to types of strains or mobile genetic element backgrounds. Comparative studies conducted in Saudi Arabia and Egypt showed that they also found similar groupings in terms of resistome profile and plasmid compatibility types[16,17].

This suggests the possible usefulness of merging genomic and plasmid-typing in future studies.

Although our findings were in accordance with the global trends, the tendency of relatively high prevalence of some of the resistance markers among these regional isolates (particularly sul1 and aadA1) might possibly indicate the selective pressures of agricultural or veterinary antibiotic use, which has been reported by recent surveys in the region[18].

These factors could be dissected further with additional research that utilize metadata involved in source and usage pattern.

Despite the promising insights provided by our machine learning–based predictions, several limitations should be acknowledged. First, hyperparameter tuning, cross-validation strategy, and feature selection procedures were not fully optimized or described in detail, which may influence model stability. Second, the models were not tested on an independent dataset, and no external or cross-dataset validation was performed. Without such validation, it remains uncertain whether the predictive performance observed here would generalize to other E. coli isolates or to different geographic regions.

Third, the Random Forest classifiers were trained on only 31 isolates, which is a relatively small sample size for machine learning.

Small datasets increase the risk of overfitting, reduce generalizability, and make feature importance estimates less robust.

These constraints highlight the need for future studies with larger cohorts and external validation to strengthen the predictive framework and ensure broader applicability.

CONCLUSIONS

The present study gives a detailed description of the genomic predictors of antibiotic resistance of 31 E. coli isolates of Iran, Turkey, and Iraq. Whole-genome analysis, chromosomal mutational profile, and machine-learning model combinations have been used to discover major resistance genes, in particular, blaTEM-1B, sul1, aadA1, and qnrS1, to have important roles in phenotypic resistance to most frequently used antibiotics.

In addition, the finding of high-impact mutations such as gyrA S83L and parC S80I mutations supported the importance of chromosomal mutations in the resistance to the fluoroquinolones. The network and cluster analytics visualization maps demonstrated that there were particular schemes of the co-occurrence of genes and groups of clusters in some areas, which underscored the influence of geographical and ecological problems. Altogether, these findings indicate the complexity of AMR evolution and prefer the use of the region-specific surveillance mechanisms.

These findings can be utilized by the research community to add additional data and plasmid typing and functional verification to better inform the community about the way forward with the idea of containing the spread of multidrug resistant E. coli with an aim of controlling and reducing the jurisdiction of the epidemic.

Acknowledgements

The authors thank University of Mosul for documenting this work.

Funding

Self-supporting

Conflict of Interest

The authors declare no conflicts of interest regarding this work.

Ethical Consideration

Not applicable

Author’s Contribution

Yilmaz Kaya: Supervision; validation; resources; project administration; writing – review & editing; and funding acquisition.

Ali Dawood: Conceptualization; Data curation; formal analysis; investigation; methodology; software; visualization; and writing – original draft.

REFERENCES

1. Holmes AH, Moore LSP, Sundsfjord A, Steinbakk M, Regmi S, Karkey A, et al. Understanding the mechanisms and drivers of antimicrobial resistance. Lancet. 2016;387(10014):176–87. DOI: 10.1016/S0140-6736(15)00473-0
2. World Health Organization. Global action plan on antimicrobial resistance. Geneva: WHO Press; 2015.
3. Tacconelli E, Carrara E, Savoldi A, Harbarth S, Mendelson M, Monnet DL, et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. Lancet Infect Dis. 2018;18(3):318–27. DOI: 10.1016/S1473-3099(17)30753-3
4. Kaper JB, Nataro JP, Mobley HLT. Pathogenic Escherichia coli. Nat Rev Microbiol. 2004;2(2):123–40. DOI: 10.1038/nrmicro818
5. Clermont O, Bonacorsi S, Bingen E. Clonal assignment and virulence gene content of Escherichia coli strains from the French population. Environ Microbiol. 2011;13(6):1615–25. DOI: 10.1111/j.1462-2920.2011.02466.x
6. Carattoli A. Plasmids and the spread of resistance. Clin Microbiol Infect. 2013;19(6):476–84. DOI: 10.1016/j.ijmm.2013.02.001
7. Poirel L, Cattoir V, Nordmann P. Plasmid-mediated quinolone resistance; interactions between human, animal, and environmental reservoirs. Antimicrob Agents Chemother. 2004;48(10):3656–61. DOI: 10.1128/AAC.48.10.3656-3661.2004
8. Cantón R, González-Alba JM, Galán JC. CTX-M enzymes: origin and diffusion. Clin Microbiol Rev. 2012;25(3):498–516. DOI: 10.1128/CMR.00070-11
9. Toleman MA, Walsh TR, Comas I, Gascuel O. Pervasive horizontal gene transfer drives the evolution of complex traits. Nat Rev Microbiol. 2006;4(4):272–82. DOI: 10.1038/nrmicro1384
10. Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile genetic elements associated with antimicrobial resistance. Microbiol Spectr. 2015;3(1):PLAS-0038-2014. DOI: 10.1128/microbiolspec.PLAS-0038-2014
11. Ruiz J, Gómez J, Navia MM, Rivera A, Trujillo M, Alberti S, et al. High levels of resistance to fluoroquinolones in clinical isolates of Escherichia coli. J Antimicrob Chemother. 2003;51(5):1109–13. DOI: 10.1093/jac/dkg190
12. Redgrave LS, Sutton SB, Webber MA, Piddock LJV. Fluoroquinolone resistance: mechanisms, impact on bacteria, and role in evolutionary success. Antimicrob Agents Chemother. 2014;58(7):3752–60. DOI: 10.1128/AAC.00013-14
13. Hu Y, Zhu Y, Ma Y, Liu F, Lu N, Yang X, et al. Gene network analysis reveals co-selection of resistance genes in microbial communities. Front Microbiol. 2019;10:435. DOI: 10.3389/fmicb.2019.00435
14. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. Transforming clinical microbiology with bacterial genome sequencing. Nat Rev Genet. 2012;13(9):601–12. DOI: 10.1038/nrg3226
15. Pesesky MW, Hussain T, Wallace M, Patel S, Andleeb S, Burnham C, et al. Evaluation of machine learning and rule-based models for predicting antibiotic resistance profiles in E. coli. Antimicrob Agents Chemother. 2016;60(9):5573–80. DOI: 10.1128/AAC.00264-16
16. Zhao X, Yu Z, Xu Z. Study of antibiotic resistance based on whole-genome sequencing data and machine learning. Front Microbiol. 2020;11:307. DOI: 10.3389/fmicb.2020.00307
17. Al-Kadmy IMS, Ali Z, Salman IM, Hadi HA, Aziz SN, Khazaal SS. Prevalence of ESBL-producing and quinolone-resistant E. coli in Iraq. Gene Rep. 2021;23:101065. DOI: 10.1016/j.genrep.2021.101065
18. Irajian G, Mirsalehian A, Sadeghi MR, Taherikalani M, Jalalifar S, Fooladi AA, et al. Resistance patterns of uropathogenic E. coli in Iran. Iran J Microbiol. 2015;7(4):193–8.
19. Başaran-Kanlıkaya H, Öztürk D, Gürler B, Onlen Y, Öngen B, Altun M, et al. Prevalence of ESBL and fluoroquinolone resistance in Turkish E. coli isolates. Microb Drug Resist. 2018;24(2):122–9. DOI: 10.1089/mdr.2017.0083
20. Berrazeg M, Diene SM, Medjahed L, Parola P, Drissi M, Raoult D, et al. Global spread and genetic diversification of E. coli resistance. Antimicrob Agents Chemother. 2014;58(8):4267–72. DOI: 10.1128/AAC.02580-14
21. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012;67(11):2640–4. DOI: 10.1093/jac/dks261
22. Alqasim A, Abu Parda M, Al-Haddad A, Al-Ramadan S, Almutairi A, Alanazi M, et al. Molecular epidemiology of multidrug-resistant E. coli in Saudi Arabia. Int J Environ Res Public Health. 2021;18(15):8186. DOI: 10.3390/ijerph18158186