

Concordance of CARD/RGI, ResFinder, and BV-BRC Workflows for ESBL Allele Detection in 100 *Escherichia coli* Genomes

Syna Bhattacharya¹

Received August 11, 2026

Accepted September 16, 2026

Electronic access October 15, 2026

Extended-spectrum beta-lactamase (ESBL) genes confer resistance to extended-spectrum cephalosporins. Genomic analysis tools may report differing results due to distinct databases and detection rules. This study compared CARD/RGI, ResFinder, and current BV-BRC Comprehensive Genome Analysis using the same set of *Escherichia coli* genomes and the reconstructed sequence-level ESBL reference. One hundred study rows were examined. Complete assembly-level genomic FASTA files were available for 99 rows. Study row 3 remained Unresolved because a complete assembly FASTA was unavailable and was excluded from binary agreement calculations. A genome was classified as Positive only when sequence evidence supported a specific recognized ESBL allele and an intact open reading frame. CARD/RGI retained Perfect and Strict hits, ResFinder used 90% identity and 60% minimum coverage using acquired-gene output, and BV-BRC used gene-level AMR/Specialty Gene evidence. There were 19 Positive genomes, 80 Negative genomes, and 1 Unresolved genome. CARD/RGI, ResFinder, and current BV-BRC matched the reconstructed reference for all 99 eligible rows. Positive percent agreement (PPA) was 100.00% (95% exact CI, 82.35% to 100.00%), negative percent agreement (NPA) was 100.00% (95% exact CI, 95.49% to 100.00%), overall percent agreement (OPA) was 100.00% (95% exact CI, 96.34% to 100.00%), and Cohen kappa was 1.000. All 19 Positive rows had matching qualifying allele sets across the reference and three workflows. The workflows agreed under the study's defined sequence-based rules. These findings do not establish phenotypic accuracy, biological ground truth, or clinical validity.

Keywords: extended-spectrum beta-lactamase, *Escherichia coli*, blaCTX-M, CARD/RGI, ResFinder, BV-BRC, antimicrobial resistance

Introduction

Antimicrobial resistance (AMR) is a major global health problem that reduces the effectiveness of commonly used antibiotics. In 2021, an estimated 1.14 million deaths were attributable to bacterial AMR, while approximately 4.71 million deaths were associated with it¹. The burden of deaths attributable to AMR is projected to increase by 2050¹. *Escherichia coli* was among the leading bacterial contributors to this global burden in 2019². Addressing this public health problem requires reliable methods for identifying the genetic determinants that contribute to resistance in these bacteria.

Extended-spectrum beta-lactamases (ESBLs) are enzymes capable of hydrolyzing extended-spectrum cephalosporins, which are critical for treating many infections³. The CTX-M enzyme family has become especially prominent in *E. coli* over recent years⁴. In contrast, the TEM and SHV enzyme families contain both ESBL variants and non-ESBL variants^{5,6}. Because of this mixed functional composition, detecting only the *bla*TEM or *bla*SHV gene family name is

not enough to classify a genome as Positive for an ESBL^{7,8}. A defensible sequence-level classification requires consideration of the exact allele, its functional classification, and the integrity of the open reading frame^{7,8}.

Whole-genome sequencing can identify genetic determinants associated with AMR. However, the final reported result depends on both the underlying genome sequence and the computational workflow used to interpret it. Several factors may affect the reported gene set, including database content, reference sequence coverage, sequence identity requirements, resistance models, nomenclature, and output interpretation rules. For example, the Comprehensive Antibiotic Resistance Database (CARD) pairs curated resistance models with the Resistance Gene Identifier (RGI) prediction engine^{9–11}. ResFinder relies on a curated database to identify acquired resistance genes and supports genotype-based interpretation^{12–14}. AMRFinderPlus uses a curated reference gene catalog and is represented in the current Bacterial and Viral Bioinformatics Resource Center (BV-BRC) Specialty Gene evidence^{15–18}. The Pathosystems Resource Integration Center (PATRIC) is relevant to this analysis only as the historical predecessor of

¹ Chagrin Falls High School; Chagrin Falls, Ohio, United States.

BV-BRC and serves as provenance for the original project¹⁹.

Previous studies have compared AMR databases, analysis methods, and phenotype predictions. Phenotype-linked accuracy and sequence-level workflow agreement are different research questions. Mahfouz et al. evaluated CARD and ResFinder using phenotype-related endpoints²⁰. Hu et al. similarly assessed the computational prediction of AMR phenotypes²¹. Other inter-method studies have documented discordant resistance gene calls when different bioinformatics pipelines analyzed shared sequence data^{22–25}. Phenotype-linked studies evaluate whether genomic findings predict laboratory susceptibility measurements^{15,26–28}. The NCBI Prokaryotic Genome Annotation Pipeline applies curated annotation rules, which reduces the independence of a reconstructed reference that relies solely on annotation evidence²⁹. To address this gap, further investigation is needed to determine row-level and allele-level agreement under one strictly defined sequence-based classification rule.

In this study, we asked how closely three current workflows agree when they receive the same complete assemblies and are interpreted under the same explicit rule for ESBL alleles. We compared the CARD/RGI workflow, ResFinder, and current BV-BRC Comprehensive Genome Analysis with a reconstructed sequence-level reference. Our primary expectation was that agreement rates might differ among the workflows. Secondary objectives were to compare normalized qualifying allele sets, to preserve separate locus information, and to investigate six accessions that had appeared discordant historically. This comparison is strictly bounded by its sequence-level scope. It does not evaluate phenotypic resistance, clinical validity, or biological ground truth.

Methods

Research Design and Study Set

This was a retrospective, sequence-based study that evaluated concordance rather than diagnostic accuracy. The original fixed study set contained 100 *Escherichia coli* study rows, where each study row served as the analytical unit. Accession mapping was preserved for every row. The dataset was historically selected to include genomes with and without annotated beta-lactamase findings. This dataset was not randomly sampled and was not designed to estimate population prevalence. The original GenBank search string and exact historical download date were not preserved. The current reconstruction and comparator analyses were completed in August 2026. Supplementary Table S1 provides the accession, assembly, and execution status for every study row.

Assembly Inputs and Eligibility

Assembly metadata and genome packages were reviewed to

determine eligibility. A complete assembly-level genomic FASTA was recovered when available. Each comparator received the same recovered assembly for a given study row. Plasmids and additional replicons were included when they were part of the assembly. A study row was technically executable only if a complete assembly-level genomic FASTA was available. A total of 99 rows met this requirement. For study row 3, accession GCF_002853715.1, only a coding-sequence FASTA was recovered. A coding-sequence FASTA cannot establish the absence of a qualifying gene across the complete assembly. As a result, study row 3 remained Unresolved. It was not classified as Negative. This row was excluded from binary agreement denominators, and no imputation was performed for it.

Reconstructed Sequence-Level ESBL Reference

The reconstructed reference was defined before the revised comparator agreement analysis. A study row was classified as Positive only when a specific recognized ESBL allele was identified, the allele was supported by the predefined functional classification hierarchy, and the sequence evidence supported an intact open reading frame. A study row was classified as Negative only when a complete available assembly was successfully screened and no qualifying ESBL allele was identified. A row remained Unresolved when the complete assembly sequence was unavailable, the exact allele could not be determined, open reading frame integrity was uncertain, or functional ESBL classification was uncertain. The TEM and SHV enzyme families contain both ESBL and non-ESBL alleles^{4–8}. Therefore, detection of only the *bla*TEM or *bla*SHV family name did not qualify as ESBL Positive. Supplementary Table S1 contains row-level reference and eligibility information, and Supplementary Table S2 contains allele and integrity evidence. This reconstructed reference is a controlled sequence-level classification. It is not an independent phenotype, biological ground truth, infallible gold standard, or clinical resistance determination.

Comparator Workflows and Classification Rules

All three current comparators were executed before the agreement calculations. For each technically executable study row, the same recovered assembly was used for all three comparators. Each comparator was reduced to the same sequence-level classification categories. A qualifying ESBL allele produced a Positive call. A successful analysis with no qualifying ESBL allele produced a Negative call. A failed or ineligible analysis remained Unresolved. Historical PATRIC calls were not substituted for current BV-BRC results.

CARD/RGI, ResFinder, and BV-BRC were selected because they were the three named workflows in the original

fixed project and could be compared using the same recovered complete assemblies.

CARD/RGI was executed locally using RGI version 6.0.5 with CARD data version 4.0.1^{9–11}. The input type was DNA/contig, and the dataset setting was whole genome. NCBI BLAST+ version 2.16.0 and PYRODIGAL were recorded as part of the execution environment. Perfect and Strict hits were retained. Loose hits were excluded. Nudged hits were excluded under the predefined rule. JSON and tab-delimited outputs were retained. Relevant preserved evidence fields included cut-off category, best Antibiotic Resistance Ontology (ARO) hit, model type, drug class, resistance mechanism, gene family, sequence identity, reference-length coverage, predicted sequence, and coordinates. ResFinder version 4.7.2 was executed with the organism setting *E. coli*, a 90% identity threshold, and a 60% minimum coverage threshold^{12–14}. The acquired-gene analysis type was used with database version 2.6.0 at commit `eeef0aa207594fe6d51badf808473de62b28cb06`. The acquired-gene result table, `ResFinder_results.tab.txt`, controlled the classification. The phenotype-prediction output did not control the classification. Allele and location evidence was retained in the allele or reference, identity, coverage, contig, and location fields.

The current BV-BRC Comprehensive Genome Analysis workflow was used for the third comparator¹⁸. Submission occurred through the official BV-BRC command-line interface version 2.0.15 using the `p3-submit-CGA` command. Input consisted of assembled contigs submitted with organism *Escherichia coli*, taxon identifier 562, domain Bacteria, and genetic code 11. Gene-level AMR/Specialty Gene evidence, specifically `specialty-amrfinder.txt` and `specialty-rgi.txt`, controlled the ESBL classification^{15–18}. The per-antibiotic phenotype table did not control the classification. Job identifiers and raw job folders were retained. Supplementary Table S3 contains the exact settings and inclusion rules.

Agreement and Allele-Level Analysis

The binary analysis denominator included only the 99 rows with a Positive or Negative reconstructed reference and completed current comparator outputs. Study row 3 was excluded because it remained Unresolved. Positive percent agreement (PPA) was defined as the proportion of reference-Positive rows classified as Positive by the tool. Negative percent agreement (NPA) was defined as the proportion of reference-Negative rows classified as Negative by the tool. Overall percent agreement (OPA) was defined as the proportion of all eligible rows with matching reference and tool classifications. Two-sided 95% Clopper-Pearson exact binomial confidence inter-

vals were used for PPA, NPA, and OPA. Cohen's kappa was calculated as a secondary chance-corrected agreement estimate. The same 99 eligible rows were used for pairwise inter-comparator agreement. Exact McNemar testing was considered for paired discordance. Because both discordant cells were zero, the test was not informative. For the allele-level analysis, qualifying allele names were normalized before comparison. Exact allele-set concordance required equality of the qualifying allele sets. Locus counts were preserved separately. Two physical loci were not collapsed into one merely because they contained the same allele.

Historical Provenance and Data Availability

Historical CARD, ResFinder, AMRFinderPlus, and PATRIC calls were preserved as provenance only. These historical calls were not used in current agreement calculations, and historical PATRIC calls were not substituted for current BV-BRC results. The six-accession audit preserved the relationship between historical and current findings without overwriting historical data. The supplementary workbook contains accessions, assembly identifiers, reconstructed-reference classifications, current comparator classifications, qualifying alleles, execution status, and unresolved reasons. The archive retained the raw-output inventory paths, exact versions, job identifiers, gene-level evidence, and SHA-256 checksums. Public sequence accessions remain available through NCBI.

Ethical Considerations

This study used publicly available bacterial genome sequence data. It did not involve human participants, identifiable private information, or vertebrate animals. Institutional review board approval was not applicable.

Results

Reference Distribution and Eligibility

The original fixed study set contained 19 reconstructed-reference Positive rows, 80 reconstructed-reference Negative rows, and one Unresolved row. Study row 3, corresponding to accession GCF_002853715.1, was the single Unresolved row. The complete assembly-level genomic FASTA was unavailable for this accession. Only a coding-sequence FASTA was recovered, which could not establish the absence of a qualifying allele across the complete assembly. Therefore, the coding-sequence FASTA was not substituted for the missing complete assembly. Study row 3 was excluded without imputation. Consequently, the binary denominator was 99 rather than 100.

Three-Workflow Agreement

The three current workflows were CARD/RGI, ResFinder, and current BV-BRC Comprehensive Genome Analysis. Each workflow produced the same binary classification as the reconstructed sequence-level ESBL reference for every eligible row. For each workflow, the counts were 19 tool-positive/reference-positive, 0 tool-positive/reference-negative, 0 tool-negative/reference-positive, and 80 tool-negative/reference-negative rows. Positive percent agreement was 100.00% (95% exact CI, 82.35% to 100.00%), negative percent agreement was 100.00% (95% exact CI, 95.49% to 100.00%), and overall percent agreement was 100.00% (95% exact CI, 96.34% to 100.00%). Cohen's kappa was 1.000, and exact McNemar testing was not informative because both discordant cells were zero. Every inter-comparator pair agreed on 99 of 99 eligible rows with a pairwise kappa of 1.000. The positive percent agreement confidence interval was wider because its denominator was only 19. These findings are presented in Table 1 and Figure 1.

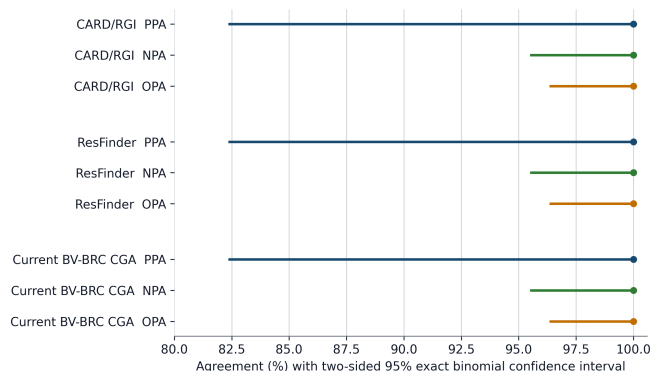


Figure. 1 Agreement estimates and confidence intervals. Each point estimate was 100.00%. Intervals are two-sided 95% Clopper-Pearson exact confidence intervals. The binary denominator was 99. Study row 3 was excluded without imputation. The reconstructed reference is sequence-based.

Six-Accession Historical Reconciliation

The six accessions that appeared discordant in the historical analysis were study rows 1, 3, 9, 11, 22, and 50. The recovered assemblies and the predefined allele-classification rule were used for reconciliation. Study rows 1, 9, 11, 22, and 50 were executable and were classified as Negative across the reconstructed reference and all three current workflows. None of these five executable rows remained discordant. Study row 3 remained Unresolved because the complete assembly was unavailable. These results are detailed in Figure 2 and Supplementary Table S4.

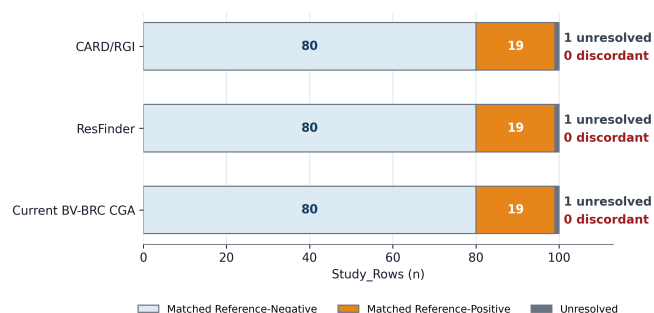


Figure. 2 Row-matched workflow outcomes. The 80 Negative and 19 Positive rows were binary matches. The one Unresolved row was study row 3. Row 3 was excluded from the 99-row binary denominator. No eligible row was discordant. The figure reports row-matched outcomes, not diagnostic accuracy

Allele-Level Concordance

All 19 reference-Positive rows had exact normalized qualifying allele-set concordance. This concordance covered the reconstructed reference, CARD/RGI, ResFinder, and current BV-BRC. The 19 genomes produced 20 distinct allele-genome observations. The specific counts were *bla*CTX-M-15 in eight genomes, *bla*CTX-M-55 in five genomes, *bla*CTX-M-27 in three genomes, *bla*CTX-M-14 in two genomes, *bla*CTX-M-1 in one genome, and *bla*CTX-M-64 in one genome. The 20 allele-genome observations came from 19 Positive genomes because study row 67 contained both *bla*CTX-M-55 and *bla*CTX-M-27. These two alleles were at distinct loci and appeared in all three tool outputs. Furthermore, current BV-BRC reported 21 qualifying loci. This occurred because study row 84 contained two separate *bla*CTX-M-15 loci, resulting in one Positive genome but two distinct BV-BRC loci. These findings are summarized in Table 2 and Figure 3.

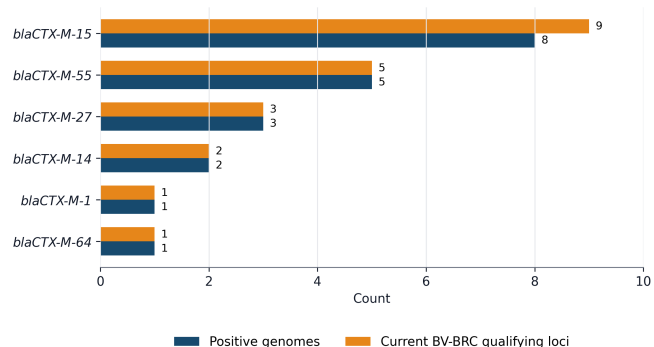


Figure. 3 Allele-genome and BV-BRC locus counts. Study row 84 produced the only difference between the two series.

Table 1 Revised agreement of current genomic workflows with the reconstructed sequence-level ESBL reference among 99 binary-eligible genomes.

Workflow	Counts*	PPA (95% CI)	NPA (95% CI)	OPA (95% CI)	Cohen's kappa
CARD/RGI	19 / 0 / 0 / 80	100.00% (82.35% to 100.00%)	100.00% (95.49% to 100.00%)	100.00% (96.34% to 100.00%)	1.000
ResFinder	19 / 0 / 0 / 80	100.00% (82.35% to 100.00%)	100.00% (95.49% to 100.00%)	100.00% (96.34% to 100.00%)	1.000
Current BV-BRC CGA	19 / 0 / 0 / 80	100.00% (82.35% to 100.00%)	100.00% (95.49% to 100.00%)	100.00% (96.34% to 100.00%)	1.000

*Counts are ordered tool-positive/reference-positive, tool-positive/reference-negative, tool-negative/reference-positive, and tool-negative/reference-negative. PPA, positive percent agreement; NPA, negative percent agreement; OPA, overall percent agreement; CI, confidence interval. Study row 3 was excluded because it remained Unresolved. These are agreement measures, not sensitivity, specificity, or diagnostic accuracy.

Table 2 Qualifying ESBL allele distribution.

Allele	Positive genomes	Study rows	Current BV-BRC loci
<i>bla</i> CTX-M-15	8	71, 73, 76, 78, 80, 83, 84, 94	9
<i>bla</i> CTX-M-55	5	67, 69, 70, 72, 77	5
<i>bla</i> CTX-M-27	3	67, 89, 97	3
<i>bla</i> CTX-M-14	2	88, 92	2
<i>bla</i> CTX-M-1	1	91	1
<i>bla</i> CTX-M-64	1	98	1

Positive-genome counts are allele-genome observations. Study row 67 contributes one observation to both *bla*CTX-M-55 and *bla*CTX-M-27. BV-BRC locus counts preserve physically separate loci. Study row 84 contributes two *bla*CTX-M-15 loci but only one Positive genome.

Discussion

Principal Findings

The principal finding of this study is that CARD/RGI, ResFinder, and current BV-BRC Comprehensive Genome Analysis matched the reconstructed sequence-level ESBL reference for all 99 evaluable genomes. Furthermore, all 19 reference-Positive rows demonstrated exact normalized qualifying allele-set concordance across all three workflows. The original expectation for this study was that agreement rates might differ among the workflows. However, that expectation was not supported in this fixed dataset, as the expected workflow differences were not observed. The observed positive percent agreement was 100.00%, with a lower 95% exact confidence limit of 82.35%. This wider confidence interval directly reflects the smaller denominator of 19 reference-Positive genomes. These findings support reproducible concordance among the workflows only under the specified dataset, versions, inputs, thresholds, and interpretation rules.

Interpretation of Workflow Concordance

The workflow concordance demonstrated here is more informative than merely comparing the aggregate total of Positive and Negative calls. Aggregate totals can match even if individual study rows are classified differently by each tool. By directly comparing row identities, this analysis confirmed that no row-level disagreements occurred among

the 99 eligible rows. Moreover, the normalized qualifying allele sets were directly compared. All 19 reference-Positive rows showed exact allele-set concordance. Study row 67 contained two different qualifying alleles, *bla*CTX-M-55 and *bla*CTX-M-27, at distinct loci. Study row 84 contained two separate loci of the same *bla*CTX-M-15 allele. These specific cases explain why allele counts, genome counts, and locus counts must be preserved separately and cannot be treated as interchangeable when interpreting concordance.

Full concordance across these workflows does not establish complete evidence independence. All classifications were derived entirely from sequence-based evidence, and the same predefined functional ESBL hierarchy was applied to interpret qualifying evidence. Although the workflows differ, their underlying databases or evidence engines are not completely independent. For instance, the current BV-BRC Specialty Gene output may include evidence derived from AMRFinderPlus and CARD/RGI^{15–18}. Previous inter-method studies have reported discordant resistance-gene calls when evaluating other databases, sequence inputs, and workflow implementations^{20,22–25}. Therefore, these current data solely support sequence-level concordance under the specified workflow conditions. This study did not measure phenotypic susceptibility, and it does not establish clinical utility.

Historical Provenance and Reconstruction

The current reconstructed analysis differs from the historical manuscript. Historically, the reconstructed-reference distribu-

tion was reported as 39 Positive and 61 Negative genomes, alongside a historical PATRIC pattern of 100 of 100 Positive classifications. These historical findings are retained only as provenance. Because the historical field-level outputs and the binary interpretation rule were not recovered, historical PATRIC results were not substituted for current BV-BRC results. The former explanation that PATRIC produced false positives because it used a broader approach was unsupported and consequently removed. The original analysis did not isolate database content, threshold effects, output type, or interpretation rules. Instead, the current BV-BRC Comprehensive Genome Analysis was newly executed using the official command-line interface, and current classifications relied on gene-level AMR/Specialty Gene evidence. When re-evaluated, historical audit rows 1, 9, 11, 22, and 50 were Negative across the reconstructed reference and all three current workflows. Study row 3 remained Unresolved and did not contribute a binary agreement cell.

Scope, Limitations, and Future Work

Several limitations must be considered. First, no matched phenotype data, antimicrobial susceptibility testing, or minimum inhibitory concentration data were available. Sequence-level ESBL status cannot establish phenotypic resistance. Second, the fixed 100-genome set was not randomly sampled. The distribution of 19 Positive, 80 Negative, and one Unresolved genome is not a measure of population prevalence, and the findings cannot automatically be generalized to broader *E. coli* populations. Third, historical retrieval information is incomplete because the original search string and exact historical download date were not preserved, although current accessions, assemblies, raw outputs, and execution records are available. Fourth, study row 3 lacked a complete genomic FASTA. This row remained Unresolved, was excluded from binary calculations, and no imputation was performed for it. Fifth, there is evidence dependence across the workflows due to the shared predefined functional hierarchy and partially related databases or evidence engines, meaning the comparisons are not fully independent. Finally, the zero observed discordance made paired difference testing uninformative. Zero discordance does not prove equivalence, as results may change with different datasets, versions, inputs, thresholds, or interpretation rules.

Future work should evaluate these workflows using larger, independently sampled collections. Future studies should match genomic findings with antimicrobial susceptibility results or minimum inhibitory concentration measurements. Additional species or epidemiological settings should also be investigated. Researchers should test different identity thresholds, coverage thresholds, and hit-category settings to separate database effects from workflow implementation ef-

fects. Comparing database versions and evaluating how different sequencing and assembly strategies, such as short-read, long-read, and hybrid assemblies, affect results would also be valuable because assembly representation can alter resistance-gene recovery^{22,23,25}. Performing this independent, phenotype-linked validation is essential to properly evaluate clinical validity and generalizability.

The three evaluated workflows agreed with the reconstructed reference for all 99 eligible genomes. Every reference-Positive row had exact normalized qualifying allele-set concordance. The six-accession historical concern was reconciled, and study row 3 appropriately remained Unresolved. Row 67 preserved two different qualifying alleles, while row 84 preserved two loci of the same allele. Ultimately, these findings apply only to the specified sequence-based conditions. Phenotypic accuracy and clinical validity were not established.

References

- 1 GBD 2021 Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. *The Lancet*. Vol. 404, pg. 1199–1226, 2024. [https://doi.org/10.1016/S0140-6736\(24\)01867-1](https://doi.org/10.1016/S0140-6736(24)01867-1).
- 2 Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet*. Vol. 399, pg. 629–655, 2022. [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0).
- 3 J. D. Pitout, K. B. Laupland. Extended-spectrum β -lactamase-producing enterobacteriaceae: an emerging public-health concern. *The Lancet Infectious Diseases*. Vol. 8, pg. 159–166, 2008. [https://doi.org/10.1016/S1473-3099\(08\)70041-0](https://doi.org/10.1016/S1473-3099(08)70041-0).
- 4 R. Cantón, J. M. González-Alba, J. C. Galán. CTX-M enzymes: origin and diffusion. *Frontiers in Microbiology*. Vol. 3, pg. 110, 2012. <https://doi.org/10.3389/fmicb.2012.00110>.
- 5 K. Bush, G. A. Jacoby. Updated functional classification of β -lactamases. *Antimicrobial Agents and Chemotherapy*. Vol. 54, pg. 969–976, 2010. <https://doi.org/10.1128/AAC.01009-09>.
- 6 K. Bush. Past and present perspectives on β -lactamases. *Antimicrobial Agents and Chemotherapy*. Vol. 62, pg. e01076-18, 2018. <https://doi.org/10.1128/AAC.01076-18>.
- 7 P. A. Bradford, R. A. Bonomo, K. Bush, A. Carattoli, M. Feldgarden, D. H. Haft, Y. Ishii, G. A. Jacoby, W. Klimke, T. Palzkill, L. Poirel, G. M. Rossolini, P. D. Tamma, C. A. Arias. Consensus on β -lactamase nomenclature. *Antimicrobial Agents and Chemotherapy*. Vol. 66, pg. e00333-22, 2022. <https://doi.org/10.1128/aac.00333-22>.
- 8 T. Naas, S. Oueslati, R. A. Bonnin, M. L. Dabos, A. Zavala, L. Dortet, P. Retailleau, B. I. Iorga. Beta-lactamase database (BLDB) - structure and function. *Journal of Enzyme Inhibition and Medicinal Chemistry*. Vol. 32, pg. 917–919, 2017. <https://doi.org/10.1080/14756366.2017.1344235>.
- 9 B. P. Alcock, A. R. Raphenya, T. T. Y. Lau, K. K. Tsang, M. Bouchard, A. Edalatmand, W. Huynh, A. L. V. Nguyen, A. A. Cheng, S. Liu, S. Y. Min, A. Miroshnichenko, H. K. Tran, R. E. Werfalli, J. A. Nasir, M. Oloni, D. J. Speicher, A. Florescu, B. Singh, M. Faltyn, A. Hernandez-Koutoucheva, A. N. Sharma, E. Bordeleau, A. C. Pawlowski, H. L. Zubyk, D. Dooley, E. Griffiths, F. Maguire, G. L. Winsor, R. G. Beiko, F. S. L. Brinkman, W. W. L. Hsiao, G. Van Domselaar, A. G. McArthur. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resis-

- tance database. *Nucleic Acids Research*. Vol. 48, pg. D517–D525, 2020. <https://doi.org/10.1093/nar/gkz935>.
- 10 B. P. Alcock, W. Huynh, R. Chalil, K. W. Smith, A. R. Raphenya, M. A. Wlodarski, A. Edalatmand, A. Petkau, S. A. Syed, K. K. Tsang, S. J. C. Baker, M. Dave, M. C. McCarthy, K. M. Mukiri, J. A. Nasir, B. Golbon, H. Imtiaz, X. Jiang, K. Kaur, M. Kwong, Z. C. Liang, K. C. Niu, P. Shan, J. Y. J. Yang, K. L. Gray, G. R. Hoad, B. Jia, T. Bhandu, L. A. Carfrae, M. A. Farha, S. French, R. Gordzevich, K. Rachwalski, M. M. Tu, E. Bordeleau, D. Dooley, E. Griffiths, H. L. Zubyk, E. D. Brown, F. Maguire, R. G. Beiko, W. W. L. Hsiao, F. S. L. Brinkman, G. Van Domseelaar, A. G. McArthur. CARD 2023: expanded curation, support for machine learning, and resistance prediction at the comprehensive antibiotic resistance database. *Nucleic Acids Research*. Vol. 51, pg. D690–D699, 2023. <https://doi.org/10.1093/nar/gkac920>.
- 11 B. Jia, A. R. Raphenya, B. Alcock, N. Waglechner, P. Guo, K. K. Tsang, B. A. Lago, B. M. Dave, S. Pereira, A. N. Sharma, S. Doshi, M. Courtot, R. Lo, L. E. Williams, J. G. Frye, T. Elsayegh, D. Sardar, E. L. Westman, A. C. Pawlowski, T. A. Johnson, F. S. L. Brinkman, G. D. Wright, A. G. McArthur. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Research*. Vol. 45, pg. D566–D573, 2017. <https://doi.org/10.1093/nar/gkw1004>.
- 12 E. Zankari, H. Hasman, S. Cosentino, M. Vestergaard, S. Rasmussen, O. Lund, F. M. Aarestrup, M. V. Larsen. Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial Chemotherapy*. Vol. 67, pg. 2640–2644, 2012. <https://doi.org/10.1093/jac/dks261>.
- 13 V. Bortolaia, R. S. Kaas, E. Ruppe, M. C. Roberts, S. Schwarz, V. Cattoir, A. Philippon, R. L. Allesoe, A. R. Rebelo, A. F. Florensa, L. Fagelhauer, T. Chakraborty, B. Neumann, G. Werner, J. K. Bender, K. Stingl, M. Nguyen, J. Coppens, B. B. Xavier, S. Malhotra-Kumar, H. Westh, M. Pinholt, M. F. Anjum, N. A. Duggett, I. Kempf, S. Nykäsenoja, S. Olkkola, K. Wiecek, A. Amaro, L. Clemente, J. Mossong, S. Losch, C. Ragimbeau, O. Lund, F. M. Aarestrup. ResFinder 4.0 for predictions of phenotypes from genotypes. *Journal of Antimicrobial Chemotherapy*. Vol. 75, pg. 3491–3500, 2020. <https://doi.org/10.1093/jac/dkaa345>.
- 14 A. F. Florensa, R. S. Kaas, P. T. L. C. Clausen, D. Aytan-Aktug, F. M. Aarestrup. ResFinder - an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microbial Genomics*. Vol. 8, pg. 000748, 2022. <https://doi.org/10.1099/mgen.0.000748>.
- 15 M. Feldgarden, V. Brover, D. H. Haft, A. B. Prasad, D. J. Slotta, I. Tolstoy, G. H. Tyson, S. Zhao, C. H. Hsu, P. F. McDermott, D. A. Tadesse, C. Morales, M. Simmons, G. Tillman, J. Wasilenko, J. P. Folster, W. Klimke. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrobial Agents and Chemotherapy*. Vol. 63, pg. e00483-19, 2019. <https://doi.org/10.1128/AAC.00483-19>.
- 16 M. Feldgarden, V. Brover, N. Gonzalez-Escalona, J. G. Frye, J. Haendiges, D. H. Haft, M. Hoffmann, J. B. Pettengill, A. B. Prasad, G. E. Tillman, G. H. Tyson, W. Klimke. AMRFinderPlus and the reference gene catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Scientific Reports*. Vol. 11, pg. 12728, 2021. <https://doi.org/10.1038/s41598-021-91456-0>.
- 17 M. Feldgarden, V. Brover, B. Fedorov, D. H. Haft, A. B. Prasad, W. Klimke. Curation of the AMRFinderPlus databases: applications, functionality and impact. *Microbial Genomics*. Vol. 8, pg. 000832, 2022. <https://doi.org/10.1099/mgen.0.000832>.
- 18 R. D. Olson, R. Assaf, T. Brettin, N. Conrad, C. Cucinell, J. J. Davis, D. M. Dempsey, A. Dickerman, E. M. Dietrich, R. W. Kenyon, M. Kuscuglu, E. J. Lefkowitz, J. Lu, D. Machi, C. Macken, C. Mao, A. Niewiadomska, M. Nguyen, G. J. Olsen, J. C. Overbeek, B. Parrello, V. Parrello, J. S. Porter, G. D. Pusch, M. Shukla, I. Singh, L. Stewart, G. Tan, C. Thomas, M. VanOeffelen, V. Vonstein, Z. S. Wallace, A. S. Warren, A. R. Wattam, F. Xia, H. Yoo, Y. Zhang, C. M. Zmasek, R. H. Scheuermann, R. L. Stevens. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Research*. Vol. 51, pg. D678–D689, 2023. <https://doi.org/10.1093/nar/gkac1003>.
- 19 A. R. Wattam, J. J. Davis, R. Assaf, S. Boisvert, T. Brettin, C. Bun, N. Conrad, E. Dietrich, T. Disz, J. L. Gabbard, S. Gerdes, C. S. Henry, R. W. Kenyon, D. Machi, C. Mao, E. K. Nordberg, G. J. Olsen, D. E. Murphy-Olson, R. Olson, R. Overbeek, B. Parrello, G. D. Pusch, M. Shukla, V. Vonstein, A. Warren, F. Xia, H. Yoo, R. L. Stevens. Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Research*. Vol. 45, pg. D535–D542, 2017. <https://doi.org/10.1093/nar/gkw1017>.
- 20 N. Mahfouz, I. Ferreira, S. Beisken, A. von Haeseler, A. E. Posch. Large-scale assessment of antimicrobial resistance marker databases for genetic phenotype prediction: a systematic review. *Journal of Antimicrobial Chemotherapy*. Vol. 75, pg. 3099–3108, 2020. <https://doi.org/10.1093/jac/dkaa257>.
- 21 K. Hu, F. Meyer, Z. L. Deng, E. Asgari, T. H. Kuo, P. C. Münch, A. C. McHardy. Assessing computational predictions of antimicrobial resistance phenotypes from microbial genomes. *Briefings in Bioinformatics*. Vol. 25, pg. bbac206, 2024. <https://doi.org/10.1093/bib/bbae206>.
- 22 T. J. Davies, J. Swann, A. E. Sheppard, H. Pickford, S. Lipworth, M. AbuOun, M. J. Ellington, P. W. Fowler, S. Hopkins, K. L. Hopkins, D. W. Crook, T. E. A. Peto, M. F. Anjum, A. S. Walker, N. Stoesser. Discordance between different bioinformatic methods for identifying resistance genes from short-read genomic data, with a focus on *Escherichia coli*. *Microbial Genomics*. Vol. 9, pg. 001151, 2023. <https://doi.org/10.1099/mgen.0.001151>.
- 23 R. M. Doyle, D. M. O'Sullivan, S. D. Aller, S. Bruchmann, T. Clark, A. Coello Pelegrin, M. Cormican, E. Diez Benavente, M. J. Ellington, E. McGrath, Y. Motro, T. Phuong Thuy Nguyen, J. Phelan, L. P. Shaw, R. A. Stabler, A. van Belkum, L. van Dorp, N. Woodford, J. Moran-Gilad, J. F. Huggett, K. A. Harris. Discordant bioinformatic predictions of antimicrobial resistance from whole-genome sequencing data of bacterial isolates: an inter-laboratory study. *Microbial Genomics*. Vol. 6, pg. 000335, 2020. <https://doi.org/10.1099/mgen.0.000335>.
- 24 P. T. L. C. Clausen, E. Zankari, F. M. Aarestrup, O. Lund. Benchmarking of methods for identification of antimicrobial resistance genes in bacterial whole genome data. *Journal of Antimicrobial Chemotherapy*. Vol. 71, pg. 2484–2488, 2016. <https://doi.org/10.1093/jac/dkw184>.
- 25 G. Maboni, R. D. P. Baptista, J. Wireman, I. Framst, A. O. Summers, S. Sanchez. Three distinct annotation platforms differ in detection of antimicrobial resistance genes in long-read, short-read, and hybrid sequences derived from total genomic DNA or from purified plasmid DNA. *Antibiotics*. Vol. 11, pg. 1400, 2022. <https://doi.org/10.3390/antibiotics11101400>.
- 26 G. H. Tyson, P. F. McDermott, C. Li, Y. Chen, D. A. Tadesse, S. Mukherjee, S. Bodeis-Jones, C. Kabera, S. A. Gaines, G. H. Loneragan, T. S. Edrington, M. Torrence, D. M. Harhay, S. Zhao. WGS accurately predicts antimicrobial resistance in *Escherichia coli*. *Journal of Antimicrobial Chemotherapy*. Vol. 70, pg. 2763–2769, 2015. <https://doi.org/10.1093/jac/dkv186>.
- 27 N. Stoesser, E. M. Batty, D. W. Eyre, M. Morgan, D. H. Wyllie, M. Del Ojo Elias, J. R. Johnson, A. S. Walker, T. E. A. Peto, D. W. Crook. Predicting antimicrobial susceptibilities for *Escherichia coli* and *Klebsiella pneumoniae* isolates using whole genomic sequence data. *Journal of Antimicrobial Chemotherapy*. Vol. 68, pg. 2234–2244, 2013. <https://doi.org/10.1093/jac/dkt180>.
- 28 E. Stubberfield, M. AbuOun, E. Sayers, H. M. O'Connor, R. M. Card, M. F. Anjum. Use of whole genome sequencing of commensal *Escherichia*

coli in pigs for antimicrobial resistance surveillance, United Kingdom, 2018. *Eurosurveillance*. Vol. 24, pg. 1900136, 2019. <https://doi.org/10.2807/1560-7917.ES.2019.24.50.1900136>.

- 29 T. Tatusova, M. DiCuccio, A. Badretdin, V. Chetvernin, E. P. Nawrocki, L. Zaslavsky, A. Lomsadze, K. D. Pruitt, M. Borodovsky, J. Ostell. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research*. Vol. 44, pg. 6614–6624, 2016. <https://doi.org/10.1093/nar/gkw569>.

Supplementary Information

The online version contains supplementary material available at <https://nhsjs.com/?p=57798>