

Tier 1: Exploratory | Comparative Genomics & Antimicrobial Resistome | Signal Rating: 3/5 Beakers | Nov 1, 2026 | 14 min read

Hierarchical Antimicrobial Resistance, Porin Disruption & Mobile HGT Island Calibration: Benchmarking Against NCBI AMRFinderPlus

Evaluating 2,253-node Reference Gene Hierarchy traversal, protein/rRNA/promoter point mutations, cefiderocol/carbapenem porin disruptions, and tetranucleotide Kullback-Leibler HGT island divergence across 10 ESKAPE benchmark cohorts

Distinguishing horizontally transferred plasmid or transposon resistance cassettes from chromosomal point mutations and outer-membrane porin disruptions is critical for computational pathogen genomics and molecular epidemiology. We benchmarked a native five-stage hierarchical antimicrobial resistance (AMR) and horizontal gene transfer (HGT) island engine against NCBI AMRFinderPlus v4.2.7 (database release 2026-08-07.1) across ten multi-species ESKAPE cohorts and official NCBI validation panels. The native engine reproduces all 72 reference resistance, virulence, stress, point-mutation, and porin-disruption calls with 100.00% gene-symbol and hierarchy-node concordance (Jaccard = 1.0000) while providing a 9.67-fold tetranucleotide Kullback-Leibler divergence separation between mobile plasmid/integron cargo and intrinsic chromosomal mutations.

1. Algorithmic & Mathematical Architecture: Coupled Reference Gene Hierarchy Traversal and Compositional HGT Island Scoring

Accurately profiling the antimicrobial resistome, virulome, and biocide/metal stress repertoire of bacterial assemblies requires resolving three distinct evolutionary mechanisms that flat sequence-similarity searches frequently conflate:

* Hierarchical Allele vs. Gene-Family Classification (fam.tsv Ancestor Climb): Within enzyme superfamilies such as class A, B, C, and D beta-lactamases (blaTEM , blaCTX-M , blaKPC , blaNDM , blaPDC , blaOXA) or phosphoethanolamine transferases (mcr-1.1), a single amino acid substitution can shift substrate specificity from narrow-spectrum penicillins to extended-spectrum cephalosporins or carbapenems (Feldgarden et al., 2019; Feldgarden et al., 2021). In the NCBI Reference Gene Hierarchy (fam.tsv , comprising 2,253 curated nodes and 10,078 reference proteins in release 2026-08-07.1), each node v defines complete-length (cov ref >= 90%) and partial-length (50% <= cov ref < 90%) identity and reference-coverage thresholds (BlastRule). Given a query alignment with amino acid identity $id\ obs = n\ ident / L$ align and reference coverage $cov\ ref = (q\ end - q\ start + 1) / L\ ref$, classification begins at the closest reference sequence's leaf node v 0 and climbs ancestor pointers (v <- parent(v)) until a reportable node satisfies $id\ obs >= id\ min(v)$ and $cov\ ref >= cov\ min(v)$. Exact 100.00% identity and 100.00% coverage on an allele node (reportable = 2 with a distinct symbol from its parent) receives method code ALLELEP (protein) or ALLELEX (translated nucleotide), whereas a 99.75% match to an allele leaf climbs to its parent gene family (for example, blaPDC-114 -> blaPDC via BLASTP / BLASTX).

* Curated Chromosomal Point Mutations & Susceptible Porin/Receptor Disruptions (POINTP/X , POINTN , POINT_DISRUPT): Clinical resistance to colistin, fluoroquinolones, macrolides, nitrofurantoin, cefiderocol, and carbapenems frequently arises not by acquiring a foreign enzyme, but via specific missense mutations in housekeeping sensors (such as *Escherichia* pmrB_C84R), non-coding rRNA/promoter substitutions (such as 23S_A2058T in the 50S peptidyl transferase center , blaTEMp_G162T , or ampC_T-14TGT), or loss-of-function frameshifts, nonsense stops, and insertion-sequence disruptions in outer-membrane entry channels (such as *Klebsiella pneumoniae* catechol siderophore receptor cirA for cefiderocol and porin ompK35 for carbapenems; AMRProt-susceptible.tsv).

* Information-Theoretic Horizontal Gene Transfer (HGT) Island Context: While reference gene cataloging identifies which resistance determinants are present, epidemiological risk assessment also depends on whether a locus is locked on the host chromosome or carried on a horizontally transmissible conjugative plasmid (IncFII , IncX3 , IncL/M), composite transposon (Tn4401 , Tn125), or Class 1 integron (intI1 - attI1 - qacEDelta1 - sul1). Our native engine augments every called element with its 256-tetranucleotide Kullback-Leibler divergence D KL (P locus ∥ Q host) (in bits) and 64-codon Relative Synonymous Codon Usage (RSCU) cosine distance against the host chromosomal baseline:

$$D_{KL} (P \text{ locus } \parallel Q \text{ host }) = \sum_{k \in \{A,C,G,T\}} \frac{P_{\text{locus}}(k)}{Q_{\text{host}}(k)} \log_2 \left(\frac{P_{\text{locus}}(k)}{Q_{\text{host}}(k)} \right)$$

To validate our five-stage native Rust implementation head-to-head against the external reference standard, we compiled the official NCBI AMRFinderPlus v4.2.7 C++ tool suite (amr_report , dna_mutation , disruption2genesymbol , and fasta_check) against the official 2026-08-07.1 database release and executed a ten-cohort calibration benchmark (Figure 1).

Generated by: Native Horizontal Gene Transfer, Tetranucleotide KL-Divergence & Resistome Scanner

2. Ten-Cohort Calibration Sweep Across Official NCBI Validation Panels & Six ESKAPE Taxgroups

We evaluated 79 protein and nucleotide sequences across ten benchmark cohorts designed to test every classification branch of the AMRFinderPlus v4.2.7 state machine:

* Official NCBI Proteome Validation Panel (COHORT-01 , *Escherichia* , 17 sequences): Tests exact allele calls (blaTEM-156 via ALLELEP), 1-mismatch ancestor family climbing (blaPDC-114 at 99.75% identity climbing to blaPDC via BLASTP ; qacR at 99.47% via BLASTP), internal vs. contig-boundary partial truncations (blaOXA-436 at 87.92% coverage -> blaOXA / blaOXA-48_fam via PARTIALP ; aph(3'')-Ib at contig 5′ boundary -> PARTIAL_CONTIG_ENDP ; sul2 at 81.55% -> PARTIALP), amino acid point mutations (pmrB_C84R and nfsA_R15C via POINTP), Shiga toxin subunits (stxA2 , stxB2 via EXACTP), and organism -specific suppression of intrinsic housekeeping *E. coli* K-12 arsenical repressor arsR (BAE77793.1).

* Official NCBI Translated Nucleotide & Promoter/ rRNA Panel (COHORT-02 , *Escherichia* , 19 contigs): Evaluates six-frame translated ALLELEX , EXACTX , BLASTX

, PARTIALX , PARTIAL_CONTIG_ENDX , premature stop detection (blaOXA-9 and blaTEM via INTERNAL_STOP), dual nonsense/missense calls (nfsA_R15C and nfsA_K141Ter via POINTX), and nucleotide-level promoter/rRNA mutations via POINTN (blaTEMP_G162T at NZ_CP095603.1 , ampC_T-14TGT 2-bp promoter insertion at NZ_CP041538.1 , and 23S_A2058T macrolide resistance at NC_004431.1).

* Official NCBI Susceptible Porin & Siderophore Disruption Panel (COHORT-03 , Klebsiella pneumoniae , 7 contigs): Evaluates multi-HSP insertion-sequence disruptions (cirA_A169insTer10 with a 3,120-bp IS insertion and cirA_N184NRHSEWTer with a 4,023-bp insertion carrying blaCTX-M-15), frameshift indels (cirA_K633LfsTer8 , cirA_Y253CfsTer5 , ompK35_R60LfsTer31), and nonsense stop codons (cirA_Q562Ter , cirA_T98Ter) conferring clinical cefiderocol and carbapenem resistance via POINT_DISRUPT .

* Six Clinical ESKAPE MDR Plasmid/Integron/SCCmec Cassettes (COHORT-04 through COHORT-09) & One 0 Housekeeping Negative Control (COHORT-10): Spans K. pneumoniae IncFII(K) Tn4401a (blaKPC-2 , blaTEM-1 , aac(6')-Ib4 , sul1 , qacEdelta1), Acinetobacter baumannii Tn125 (blaNDM-1 , ble-MBL , aph(3')-VI , armA , blaOXA-23), E. coli ST131 IncF (blaCTX-M-15 , blaOXA-1 , aac(6')-Ib-cr , tet(A) , dfrA17 , mph(A)), Staphylococcus aureus USA300 MRSA SCCmec Type IV (mecA , blaZ , erm(C) , qacA , merA), Pseudomonas aeruginosa ST235 Class 1 integron (blaVIM-2 , aac(6')-Ib4 , sul1 , qacEdelta1 , catB3), Salmonella enterica IncX3 / SGI1 (mcr-1.1 , blaSHV-12 , floR , tet(G) , aadA2), and a deliberate 0 negative control of conserved E. coli K-12 housekeeping proteins (rpoB , dnaK , groEL , recA , plus suppressed arsR_K-12).

Generated by: Native Horizontal Gene Transfer, Tetranucleotide KL-Divergence & Resistome Scanner

Across all ten cohorts, the native engine matches 72 out of 72 reference calls from AMRFinderPlus v4.2.7 with a gene-symbol Jaccard index of 1.0000 , 100.00% fam.tsv hierarchy node concordance , 100.00% method classification concordance , and 0.00% identity or coverage deviation (Table 1). Simultaneously, the compositional HGT scanner separates mobile plasmid, transposon, prophage, and integron resistance cargo (mean D KL = 0.435 bits) from chromosomal point mutations and porin disruptions (mean D KL = 0.045 bits , a 9.67-fold compositional separation).

Cohort ID	Organism Taxgroup	Assay Mode	Input Sequences	Total Length (bp/aa)	Native Calls	AMRFinderPlus Calls	Symbol Jaccard	Hierarchy Node Concordance (%)	Method Concordance (%)	Mean HGT D KL (bits)	Dominant Mobilome / Genomic Context
COHORT-01	Escherichia	Protein + GFF	17	4668	13	13	1.0000	100.00	100.00	0.356	Mixed plasmid AMR (blaTEM-156 , blaPDC , vanG), Stx2a prophage, and chromosomal pmrB / nfsA
COHORT-02	Escherichia	Nucleotide	19	31538	20	20	1.0000	100.00	100.00	0.325	6-frame ALLELEX / BLASTX / INTERNAL_STOP + POINTN (blaTEMP_G162T , ampC_T-14TGT , 23S_A2058T)
COHORT-03	Klebsiella pneumoniae	Nucleotide Disrupt	7	24486	8	8	1.0000	100.00	100.00	0.089	7 chromosomal cirA / ompK35

POINT_DISRUPT loci + 1 mobile blaCTX-M-15 insertion |

| COHORT-04 | Klebsiella pneumoniae | Protein | 5 | 1157 | 5 | 5 | 1.0000 | 100.00 | 100.00 | 0.440 | IncFII(K) conjugative plasmid + Tn4401a transposon (blaKPC-2 , blaTEM-1 , sul1 , qacEdelta1) |

| COHORT-05 | Acinetobacter baumannii | Protein | 5 | 1182 | 5 | 5 | 1.0000 | 100.00 | 100.00 | 0.485 | Tn125 / ISAba125 composite transposon island (blaNDM-1 , ble-MBL , aph(3')-VI , armA , blaOXA-23) |

| COHORT-06 | Escherichia | Protein | 6 | 1624 | 6 | 6 | 1.0000 | 100.00 | 100.00 | 0.422 | Pandemic ST131 IncF MDR plasmid cassette (blaCTX-M-15 , blaOXA-1 , aac(6')-Ib-cr , tet(A)) |

| COHORT-07 | Staphylococcus aureus | Protein | 5 | 2262 | 5 | 5 | 1.0000 | 100.00 | 100.00 | 0.418 | USA300 MRSA SCCmec Type IV + biocide/mercury island (mecA , blaZ , erm(C) , qacA , merA) |

| COHORT-08 | Pseudomonas aeruginosa | Protein | 5 | 1040 | 5 | 5 | 1.0000 | 100.00 | 100.00 | 0.428 | ST235 Class 1 integron intI1 - attI1 cassette (blaVIM-2 , aac(6')-Ib4 , sul1 , qacEdelta1 , catB3) |

| COHORT-09 | Salmonella enterica | Protein | 5 | 1882 | 5 | 5 | 1.0000 | 100.00 | 100.00 | 0.442 | IncX3 / ISAp1 & SGI1 island (mcr-1.1 , blaSHV-12 , floR , tet(G) , aadA2) |

| COHORT-10 | Escherichia | Negative Control | 5 | 2996 | 0 | 0 | 1.0000 | 100.00 | 100.00 | 0.036 | 0 Negative control: conserved rpoB , dnaK , groEL , recA + suppressed intrinsic arsR_K-12 |

Generated by: NCBI Entrez E-Utilities * CRV Science Labs PoliteCacheManager * Native Horizontal Gene Transfer, Tetranucleotide KL-Divergence & Resistome Scanner

3. Focal-Region Sequence Alignments & Representative Locus Telemetry Across Method Classes

To illustrate how the native engine and AMRFinderPlus v4.2.7 evaluate distinct resistance mechanisms at single-residue and single-nucleotide resolution, Table 2 presents 18 representative calls spanning all 10 method classifications, followed by annotated focal-region alignments (Figure 3) across three diagnostic regimes: (A) the blaTEM-156 vs. blaKPC-2 Ambler Class A active-site Ser70-X-X-Lys73 and Omega-loop region, (B) the K. pneumoniae cirA catechol siderophore receptor truncation and IS-insertion hotspot conferring cefiderocol resistance, and (C) the Escherichia 23S rRNA Domain V peptidyl transferase center (A2058T) and ampC promoter (T-14TGT) non-coding resistance mutations.

Cohort	Contig / Protein ID	Element Symbol	Hierarchy Node	Scope	Type / Subtype	Drug Class / Subclass	Method	Target Len	Ref Len	Ref Cov (%)	Identity (%)	Closest Ref Accession	HGT D KL (bits)
COHORT-01	contig01	blaTEM-156	blaTEM-156	core	AMR / AMR	BETA-LACTAM / BETA-LACTAM	ALLELEP	286	286	100.00	100.00	WP_061158039.1	0.427
COHORT-01	contig02	blaPDC	blaPDC	core	AMR / AMR								

BETA-LACTAM / CEPHALOSPORIN || BLASTP || 397 || 397 || 100.00 || 99.75 ||
 WP_061189306.1 || 0.448 |

| COHORT-01 || contig03 || blaOXA || blaOXA-48_fam || core || AMR / AMR ||
 BETA-LACTAM / BETA-LACTAM || PARTIALP || 233 || 265 || 87.92 || 100.00 ||
 WP_058842180.1 || 0.440 |

| COHORT-01 || contig09 || aph(3'')-Ib || aph(3'')-Ib || core || AMR / AMR ||
 AMINOGLYCOSIDE / STREPTOMYCIN || PARTIAL_CONTIG_ENDP || 225 || 267 || 81.27 ||
 100.00 || WP_001082319.1 || 0.441 |

| COHORT-01 || contig14 || pmrB_C84R || pmrB || core || AMR / POINT ||
 COLISTIN / COLISTIN || POINTP || 363 || 363 || 100.00 || 99.72 ||
 WP_001300761.1 || 0.059 |

| COHORT-01 || contig18 || stxA2 || stxA2_acd || plus || VIRULENCE / VIRULENCE
 || STX2 / stxA2 || EXACTP || 319 || 319 || 100.00 || 100.00 || TJA36680.1 ||
 0.444 |

| COHORT-02 || contig01 || blaTEMp_G162T || NA || core || AMR / POINT ||
 BETA-LACTAM / AMOXICILLIN-CLAVULANATE || POINTN || 984 || 1176 || 83.67 ||
 99.80 || NZ_CP095603.1 || 0.415 |

| COHORT-02 || contig10 || blaOXA || blaOXA-9_fam || core || AMR / AMR ||
 BETA-LACTAM / BETA-LACTAM || INTERNAL_STOP || 274 || 274 || 100.00 || 99.64 ||
 WP_000722315.1 || 0.422 |

| COHORT-02 || contig15 || 23S_A2058T || NA || core || AMR / POINT ||
 MACROLIDE / AZITHROMYCIN || POINTN || 2905 || 2905 || 100.00 || 99.97 ||
 NC_004431.1 || 0.041 |

| COHORT-02 || contig16 || nfsA_K141Ter || nfsA || core || AMR / POINT ||
 NITROFURAN / NITROFURANTOIN || POINTX || 240 || 240 || 100.00 || 99.17 ||
 WP_089631889.1 || 0.050 |

| COHORT-02 || contig17 || ampC_T-14TGT || NA || core || AMR / POINT ||
 BETA-LACTAM / CEPHALOSPORIN || POINTN || 247 || 245 || 100.00 || 99.19 ||
 NZ_CP041538.1 || 0.044 |

| COHORT-03 || cirA_A169insTer10 || cirA_A169insTer10 || cirA || core || AMR /
 POINT_DISRUPT || BETA-LACTAM / CEFIDEROCOL || POINTX || 1696 || 657 || 100.00 ||
 99.85 || WP_002912926.1 || 0.041 |

| COHORT-03 || ompK35_R60LfsTer31 || ompK35_R60LfsTer31 || ompK35 || core ||
 AMR / POINT_DISRUPT || BETA-LACTAM / CARBAPENEM || POINTX || 361 || 359 ||
 100.00 || 100.00 || WP_004141771.1 || 0.042 |

| COHORT-04 || blaKPC-2 || blaKPC-2 || blaKPC-2 || core || AMR / AMR ||
 BETA-LACTAM / CARBAPENEM || ALLELEP || 293 || 293 || 100.00 || 100.00 ||
 WP_004199234.1 || 0.451 |

| COHORT-05 || blaNDM-1 || blaNDM-1 || blaNDM-1 || core || AMR / AMR ||
 BETA-LACTAM / CARBAPENEM || ALLELEP || 270 || 270 || 100.00 || 100.00 ||
 WP_004201164.1 || 0.495 |

| COHORT-07 || mecA || mecA || mecA || core || AMR / AMR || BETA-LACTAM /

METHICILLIN || EXACTP || 668 || 668 || 100.00 || 100.00 || WP_000721306.1 || 0.432 |

| COHORT-08 || blaVIM-2 || blaVIM-2 || blaVIM-2 || core || AMR / AMR || BETA-LACTAM / CARBAPENEM || ALLELEP || 266 || 266 || 100.00 || 100.00 || WP_003158487.1 || 0.442 |

| COHORT-09 || mcr-1.1 || mcr-1.1 || mcr-1.1 || core || AMR / AMR || COLISTIN / COLISTIN || ALLELEP || 541 || 541 || 100.00 || 100.00 || WP_049589868.1 || 0.438 |

Generated by: Native Horizontal Gene Transfer, Tetranucleotide KL-Divergence & Resistome Scanner

Generated by: Native Secondary-Structure rRNA / ITS Compensatory Base Change (CBC) & Partitioned Systematics Engine

4. Phylogenetic Relationships Across Ambler Class A, B, C & D Beta-Lactamase Superfamilies

To contextualize the hierarchical fam.tsv node assignments across the beta-lactamase superfamily, we constructed a Neighbor-Joining cladogram and Newick phylogeny across twelve representative reference enzymes from our benchmark cohorts spanning Ambler Class A serine beta-lactamases (blaTEM-1 , blaTEM-156 , blaSHV-12 , blaKPC-2 , blaCTX-M-15 , blaZ), Class B1 metallo-beta-lactamases (blaNDM-1 , blaVIM-2), Class C AmpC cephalosporinases (blaPDC-114 , blaEC-28), and Class D OXA oxacillinases/carbapenemases (blaOXA-1 , blaOXA-23 , blaOXA-436) (Figures 4A and 4B).

Generated by: Native K2P / TN93 Distance, Neighbor-Joining & ABGD Systematics Engine
* Native Horizontal Gene Transfer, Tetranucleotide KL-Divergence & Resistome Scanner

Generated by: Native K2P / TN93 Distance, Neighbor-Joining & ABGD Systematics Engine

Discussion & Key Takeaways

1. Signal Rating Assessment (3 Comparative Synthesis): This ten-cohort benchmark spans 79 sequences and 72 verified resistance, virulence, stress, point-mutation, and porin-disruption calls across six ESKAPE taxgroups (Escherichia , Klebsiella pneumoniae , Acinetobacter baumannii , Staphylococcus aureus , Pseudomonas aeruginosa , and Salmonella enterica), paired with a deliberate 0 housekeeping negative control (COHORT-10). In accordance with our signal rating rubric, this multi-clade cross-engine synthesis earns a 3 Comparative Synthesis rating.

2. Why Multi-HSP Split Chaining Is Essential for Susceptible Porin & Siderophore Receptor Disruptions: In clinical K. pneumoniae isolates resistant to cefiderocol or carbapenems, loss of function in cirA or ompK35 is frequently driven by insertion sequences (such as IS26 or ISKpn26) jumping directly into the coding sequence (as seen in cirA_A169insTer10 and cirA_N184NRHSEWTer). Standard single-HSP local alignment drops the downstream half of the split open reading frame, misclassifying an insertion disruption as a simple partial fragment. Retaining all secondary HSPs (--max-hsps 0) and evaluating split-ORF coordinates through disruption2genesymbol recovers all eight POINT_DISRUPT and embedded blaCTX-M-15 calls on test_disrupt.fa

with 100.00% concordance.

3. Candidate Hypotheses for Downstream Genomic Surveillance: Pairing hierarchical fam.tsv gene typing with 256-tetranucleotide Kullback-Leibler divergence ($D_{KL} = 0.435$ bits on mobile plasmid/integron cassettes vs. 0.045 bits on chromosomal point mutations) provides a fast computational signal that may help molecular epidemiologists prioritize horizontally transferable carbapenemase and colistin-resistance islands during outbreak surveillance.

References & Primary Sources

- * [1] Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, Hoffmann M, Pettengill JB, Prasad AB, Tillman GE, Tyson GH, Klimke W (2021). AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Scientific Reports* 11:12728. <https://doi.org/10.1038/s41598-021-91456-0>
- * [2] Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, Tyson GH, Zhao S, Hsu CH, McDermott PF, Tadesse DA, Morales C, Simmons M, Tillman G, Wasilenko J, Folster JP, Klimke W (2019). Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrobial Agents and Chemotherapy* 63(11):e00483-19. <https://doi.org/10.1128/AAC.00483-19>
- * [3] Bush K, Jacoby GA (2010). Updated functional classification of beta-lactamases. *Antimicrobial Agents and Chemotherapy* 54(3):969-976. <https://doi.org/10.1128/AAC.01009-09>
- * [4] Poirel L, Jayol A, Nordmann P (2017). Polymyxins: antibacterial activity, susceptibility testing, and resistance mechanisms encoded by plasmids or chromosomes. *Clinical Microbiology Reviews* 30(2):557-596. <https://doi.org/10.1128/CMR.00064-16>