

Genomics & Complexity Theory | Signal Rating: 1/5 Stars | Oct 28, 2026 | 9 min read

Horizontal Gene Transfer & Mobile Resistome Baseline Scan: Tetranucleotide KL-Divergence Pinpoints a 20.2 kb Acinetobacter-Derived blaNDM-1 Integrative Conjugative Element in Pseudomonas

Executing the CRV Science Labs Rust bio-suite hgt-scan pipeline across a 90 kb Pseudomonas chromosome benchmark resolves an unameliorated 20.22 kb tRNA-Gly integrative conjugative element (GI-ICE-PAE-01) carrying ISAb₁₂₅-bracketed blaNDM-1, bleMBL, and sul1 resistance cassettes.

1. Abstract & Information-Theoretic Compositional Anomaly Principle

Bacterial chromosomes maintain a species-specific genomic signature -- an equilibrium k-mer frequency spectrum and synonymous codon usage bias enforced by host DNA replication, methylation, and mismatch-repair machinery [1]. When a clinical or environmental isolate acquires foreign DNA via a conjugative plasmid, prophage, or integrative conjugative element (ICE), the newly inserted block stands out as an information-theoretic anomaly before evolutionary amelioration gradually homogenizes its base composition toward the host background [2].

To benchmark Pipeline 3 (bio-suite hgt-scan), we executed the multi-threaded Rust worker-agent across a 90,000 bp Pseudomonas aeruginosa benchmark chromosome segment (NZ_CP089412.1_Pseudomonas_aeruginosa_ENV26, overall background GC = 60.27%, core host flank GC = 73.94%) using a 5,000 bp sliding window and 2,500 bp step size (35 windows total). For each window, the engine computes the 256-dimensional Laplace-smoothed tetranucleotide (4-mer) probability distribution and evaluates the Kullback-Leibler Divergence (D_{KL} in bits) and 64-codon Relative Synonymous Codon Usage (RSCU) cosine distance against the whole-contig background, followed by 10 kb flanking mobility boundary interrogation (tRNA-scan-SE 2.0 [3], VirSorter2 / PhiSpy, ISEScan [4], and MOB-suite) and antimicrobial resistance cargo screening (AMRFinderPlus [5] and CARD [6]).

2. Sliding-Window Tetranucleotide KL-Divergence & RSCU Telemetry

Across the 35 overlapping 5 kb windows, windows HGT-W01 through HGT-W13 and HGT-W23 through HGT-W35 reflect the canonical GC-rich Pseudomonas host backbone (GC = 73.36%-74.54%, negative GC-skew = -0.0641 to -0.0797, baseline D_{KL} = 0.3655-0.3836 bits). Between coordinates 35,000 bp and 55,000 bp (windows HGT-W15 through HGT-W21), the compositional profile undergoes a sharp phase transition: GC content drops to 12.16%-12.66%, GC-skew inverts to +0.4381 to +0.5635 (Z = +2.2240), and tetranucleotide Kullback-Leibler divergence surges more than fivefold to 1.9437-1.9759 bits (Z = +1.9640 to +2.0144) alongside an RSCU cosine distance of 0.4619-0.6902. Click Copy for Excel (TSV) to export the sliding-window compositional telemetry:

| Window ID || Start (bp) || End (bp) || GC (%) || GC Skew || GC Skew Z ||
4-mer KL Div (bits) || KL Div Z || RSCU Cosine Dist || Candidate Island |

HGT-W01	0	5000	73.94	-0.0695	0.5194	0.3671	-0.5025	0.0676	false
HGT-W07	15000	20000	73.62	-0.0649	0.4995	0.3711	-0.4962	0.0676	false
HGT-W13	30000	35000	73.88	-0.0704	0.5233	0.3693	-0.4991	0.0676	false
HGT-W14	32500	37500	43.38	0.0244	0.1124	0.3276	-0.5643	0.6377	false
HGT-W15	35000	40000	12.66	0.5166	2.0207	1.9437	1.9640	0.6902	true
HGT-W16	37500	42500	12.60	0.4381	1.6805	1.9759	2.0144	0.4619	true
HGT-W17	40000	45000	12.46	0.5409	2.1260	1.9642	1.9961	0.5850	true
HGT-W18	42500	47500	12.28	0.5635	2.2240	1.9650	1.9973	0.6811	true
HGT-W19	45000	50000	12.16	0.4967	1.9345	1.9725	2.0091	0.4624	true
HGT-W20	47500	52500	12.18	0.5107	1.9952	1.9727	2.0094	0.5835	true
HGT-W21	50000	55000	12.52	0.4952	1.9280	1.9716	2.0076	0.6766	true
HGT-W22	52500	57500	43.32	0.0092	0.1783	0.3320	-0.5574	0.0200	false
HGT-W23	55000	60000	74.04	-0.0686	0.5155	0.3655	-0.5050	0.5308	false
HGT-W30	72500	77500	73.92	-0.0682	0.5138	0.3673	-0.5022	0.4670	false
HGT-W35	85000	90000	73.60	-0.0641	0.4960	0.3718	-0.4952	0.5288	false

3. Mobility Boundary Mapping & Resistome Cargo Interrogation (GI-ICE-PAE-01)

In Step 2 and Step 3, bio-suite hgt-scan scans the 10 kb boundaries flanking the compositional anomaly (34,890-55,110 bp; 20.22 kb total) and screens internal open reading frames against AMRFinderPlus , CARD , and VFDB . The island is precisely bracketed by an 18 bp attL / attR target-site duplication at the 3' end of tRNA -Gly(GCC) alongside a XerC/IntI1 tyrosine site-specific recombinase (PF00589) and a MOB P1 conjugative relaxase, classifying GI-ICE-PAE-01 as an active integrative conjugative element:

Feature / ORF ID	Start (bp)	End (bp)	Strand	Symbol / Element
------------------	------------	----------	--------	------------------

Category / Drug Class || Database / Source || Identity (%) || E-Value |
 | BOUNDARY-01 || 34890 || 34965 || + || tRNA-Gly(GCC) attL || Mobility
 Integration Site || tRNAscan-SE 2.0 || 100.00 || 1.4e-19 |
 | BOUNDARY-02 || 35420 || 36410 || + || intI1 / XerC || Tyrosine Recombinase
 (PF00589) || VirSorter2 / PhiSpy || 98.40 || 3.2e-64 |
 | BOUNDARY-03 || 41200 || 42285 || + || ISAb125 || DDE Transposase (Tn125
 composite) || ISEScan || 99.80 || 8.5e-52 |
 | ENV26_ORF_0041 || 42450 || 43263 || + || blaNDM-1 || Carbapenems /
 Beta-Lactams || AMRFinderPlus / CARD ARO:3000589 || 99.63 || 0.0 |
 | ENV26_ORF_0042 || 43310 || 43675 || + || bleMBL || Glycopeptide / Bleomycin |
 | CARD ARO:3002819 || 100.00 || 1.1e-84 |
 | ENV26_ORF_0046 || 46800 || 47610 || + || sul1 || Sulfonamides
 (Dihydropteroate synthase) || CARD ARO:3000410 || 99.28 || 0.0 |
 | ENV26_ORF_0049 || 49200 || 50640 || + || feOAB-tonB-vir || Siderophore Iron
 Acquisition || VFDB VFG001842 || 96.40 || 4.5e-140 |
 | BOUNDARY-04 || 51800 || 53120 || - || mobP1 / traG || Conjugative Relaxase &
 T4SS Coupling || MOB-suite / PlasmidFinder || 97.10 || 2.1e-48 |
 | BOUNDARY-05 || 55110 || 55128 || + || tRNA-Gly(GCC) attR || 18 bp Direct
 Repeat Restoration || tRNAscan-SE 2.0 || 100.00 || 1.8e-18 |

4. Mechanistic Synthesis & 1* Signal Rating Assessment

The co-occurrence of an unameliorated AT-rich tetranucleotide signature (D KL = 1.9759 bits) with an upstream ISAb125 insertion sequence and downstream ble MBL gene reproduces the canonical Tn125 transposon architecture originally mobilized from *Acinetobacter baumannii* into Enterobacterales and *Pseudomonas* spp. [5, 6] . Because the island has not yet undergone mutational amelioration toward the high-GC *Pseudomonas* background, bio-suite hgt-scan flags the transfer event directly from primary sequence composition prior to phenotypic susceptibility testing.

Signal Rating Assessment (1* Baseline / Review Note): This report is rated 1* because it serves as a deterministic baseline validation of the bio-suite hgt-scan pipeline on a single 90 kb reference benchmark contig (n = 35 windows). Achieving a 3*-5* discovery rating requires executing the scanner across newly sequenced, uncharacterized clinical or environmental metagenome-assembled genomes (MAGs) and confirming novel mobile resistance elements via long-read junction spanning and conjugation assays.

References & Primary Sources

- * [1] Karlin, S., Mrzek, J., & Campbell, A. M. (1997). Compositional biases of bacterial genomes and evolutionary implications. *Journal of Bacteriology*, 179(12), 3899-3913. <https://doi.org/10.1128/jb.179.12.3899-3913.1997>
- * [2] Lawrence, J. G., & Ochman, H. (1997). Amelioration of bacterial genomes: rates of change and exchange. *Journal of Molecular Evolution*, 44(4), 383-397. <https://doi.org/10.1007/PL00006158>

* [3] Chan, P. P., Lin, B. Y., Mak, A. J., & Lowe, T. M. (2021). tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Research*, 49(16), 9077-9096. <https://doi.org/10.1093/nar/gkab688>

* [4] Xie, Z., & Tang, H. (2017). ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics*, 33(21), 3340-3347. <https://doi.org/10.1093/bioinformatics/btx433>

* [5] Feldgarden, M., Brover, V., Gonzalez-Escalona, N., et al. (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>

* [6] Alcock, B. P., Huynh, W., Chalil, R., et al. (2023). CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Research*, 51(D1), D690-D699. <https://doi.org/10.1093/nar/gkac920>