

Microbial Genomics & Clinical Diagnostics | Signal Rating: 1/5 Stars | Oct 28, 2026 | 8 min read

Intragenomic 16S rRNA Multi-Copy Operon Heterogeneity Across 21 Bacterial Operons

Complete operon-by-operon sequence comparison across 21 intragenomic 16S rRNA copies (32.4 kb) demonstrating how micro-heterogeneity in variable loops V1 and V6 inflates single-nucleotide microbiome diversity estimates.

| Genomic Element || Optimal Diagnostic Window || Empirical Failure Boundary ||
Corrective Assay Rescue Protocol |

| Prokaryotic 16S SSU rRNA (rrsA-rrsH Multi-Copy Operons) || Bacterial Phylum, Class, Order, Family & Deep Genus (d_K2P = 0.112-0.246; stem entropy H = 0.49 bits) || Congeneric & enteric sister species (E. coli vs. S. enterica ABGD gap = -0.0013; intragenomic rrsA-rrsH divergence 0.0091 > inter-species V3-V4 distance 0.0078) and 100%-identity ASV splitting (H_d = 0.8952 across 4 intragenomic haplotypes). || Full-length rrnA-rrnH operon phasing via long-read sequencing paired with single-copy housekeeping genes (rpoB, gyrB) or whole-contig tetranucleotide D_KL. |

Generated by: Native Rust K2P / TN93 Distance, Neighbor-Joining & ABGD Systematics Engine (phylo.rs) * Native Rust Haplotype Clustering & Minimum Spanning Network Engine (haplo.rs) * Native Rust Horizontal Gene Transfer, Tetranucleotide KL-Divergence & Resistome Scanner (hgt_bgc.rs)

Bacterial chromosomes carry between 1 and 15 copies of the 16S rRNA operon, and intragenomic divergence among paralogous copies within a single cell can artificially inflate species richness in zero-radius amplicon sequencing workflows. Comparing 21 complete 16S rRNA operons (32.4 kb) across *Escherichia coli* K-12 (rrnA-rrnH), *Salmonella enterica* LT2, and *Bacillus subtilis* 168 suggests that a single *E. coli* chromosome harbors 4 distinct 16S haplotypes differing by up to 11 nucleotides in variable loops V1 and V6.

Executive Summary: The prokaryotic 16S small-subunit ribosomal RNA gene (~1,542 bp) is the most widely sequenced taxonomic marker in clinical and environmental microbiology. However, bacterial chromosomes carry between 1 and 15 copies of the rrn ribosomal operon (16S - tRNA - 23S - 5S). In high-resolution Amplicon Sequence Variant (ASV) workflows that call zero-radius OTUs at 100% nucleotide identity, intragenomic sequence divergence among paralogous rrn operons within a single bacterial cell can artificially inflate species richness counts. Here we analyze 21 complete 16S rRNA operons across *Escherichia coli* K-12 MG1655 (7 operons: rrnA , rrnB , rrnC , rrnD , rrnE , rrnG , rrnH), *Salmonella enterica* serovar Typhimurium LT2 (7 operons), and *Bacillus subtilis* 168 (7 representative operons: rrnO , rrnA , rrnJ , rrnW , rrnI , rrnE , rrnB).

1. ASCII Haplotype Cluster Network Viewpoint of Intragenomic 16S Operon Paralogs

Executing genomic analysis at a 3-step parsimony threshold collapses the 21 ribosomal operons into h = 9 unique haplotypes (H01 - H09 , Nei's Haplotype

Diversity $H_d = 0.8952$, Shannon entropy $H_{hap} = 2.9464$ bits) partitioned into 3 species-level clusters:

Generated by: Native Rust Haplotype Clustering & Minimum Spanning Network Engine (haplo.rs)

As shown inside CLUSTER-02 (*Escherichia coli* K-12), a single clonal bacterial chromosome generates four distinct 16S haplotypes : the modal hub H02 (shared by *rrnA* , *rrnC* , and *rrnE* , $n = 3$), the 1-step V1-loop satellite H05 (shared by *rrnD* and *rrnH* , $n = 2$), and two 2-step satellites H07 (*rrnG*) and H08 (*rrnB*).

2. Focal-Region Sequence Alignment Summary (16S Variable Loop V1 / Helix H6) & Polymorphic Matrix

Rather than dumping the full 1,542 bp 16S rRNA alignment, the focal window below (nt 61..120) highlights the exact 60 bp stem-loop region encompassing Variable Loop V1 (Helix H6) where intragenomic *E. coli* K-12 operons *rrnD* / *rrnH* / *rrnG* diverge from *rrnA* / *rrnC* / *rrnE* while the flanking ribosomal stems remain 100% invariant (*):

Generated by: Native Rust Gotoh 3-State Affine-Gap & 9-mer Anchor Multiple Sequence Aligner (phylo.rs)

Generated by: Native Rust Haplotype Clustering & Minimum Spanning Network Engine (haplo.rs)

Haplotype ID	Cluster ID	Reference Taxon	Operon Copy Count	Relative Freq	Cluster Role	Member Chromosomal Operons	Substitutions vs Modal H01
H01	CLUSTER-01	<i>Salmonella enterica</i> LT2	5	0.2381	* Hub	NC_003197.2_rrnA, NC_003197.2_rrnB, NC_003197.2_rrnC, NC_003197.2_rrnD, NC_003197.2_rrnE	Modal Reference (H01)
H02	CLUSTER-02	<i>Bacillus subtilis</i> 168	4	0.1905	* Hub	NC_000964.3_rrnO, NC_000964.3_rrnA, NC_000964.3_rrnJ, NC_000964.3_rrnW	[object Object], [object Object], [object Object], [object Object], [object Object], [object Object], [object Object], [object Object], [object Object], [object Object] (+18 more)
H03	CLUSTER-03	<i>Escherichia coli</i> K-12	3	0.1429	* Hub	NC_000913.3_rrnA, NC_000913.3_rrnC, NC_000913.3_rrnE	[object Object], [object Object], [object Object], [object Object], [object Object], [object Object], [object Object], [object Object]
H04	CLUSTER-03	<i>Escherichia coli</i> K-12	2	0.0952	Paralog Satellite	NC_000913.3_rrnD, NC_000913.3_rrnH	[object Object], [object Object], [object Object], [object Object], [object Object], [object Object], [object Object], [object Object]
H05	CLUSTER-02	<i>Bacillus subtilis</i> 168	2	0.0952	Paralog Satellite	NC_000964.3_rrnI, NC_000964.3_rrnE	[object Object], [object Object], [object Object], [object Object], [object Object], [object Object], [object Object], [object Object] (+19 more)

| H06 | | CLUSTER-01 | | Salmonella enterica LT2 | | 2 | | 0.0952 | | Paralog
Satellite | | NC_003197.2_rrnG, NC_003197.2_rrnH | | [object Object] |

| H07 | | CLUSTER-03 | | Escherichia coli K-12 | | 1 | | 0.0476 | | Paralog
Satellite | | NC_000913.3_rrnB | | [object Object], [object Object], [object
Object], [object Object], [object Object], [object Object], [object
Object], [object Object], [object Object] |

| H08 | | CLUSTER-03 | | Escherichia coli K-12 | | 1 | | 0.0476 | | Paralog
Satellite | | NC_000913.3_rrnG | | [object Object], [object Object], [object
Object], [object Object], [object Object], [object Object], [object Object], [object
Object], [object Object], [object Object] |

| H09 | | CLUSTER-02 | | Bacillus subtilis 168 | | 1 | | 0.0476 | | Paralog
Satellite | | NC_000964.3_rrnB | | [object Object], [object Object], [object
Object], [object Object], [object Object], [object Object], [object Object], [object
Object], [object Object], [object Object] (+20 more) |

Generated by: Native Rust Haplotype Clustering & Minimum Spanning Network Engine (haplo.rs)

Discussion & Key Takeaways

When microbiome tests treat every single-letter DNA difference as a separate bacterial species, a single microbe carrying slightly different copies of its own 16S gene gets counted multiple times. In common gut bacteria like *Escherichia coli*, the seven 16S gene copies inside a single cell split into four distinct sequence variants.

The practical takeaway for clinical and environmental diagnostics is straightforward: grouping gene copies that differ by only a few nucleotides within variable loops V1 and V6 reconciles intragenomic paralogs back to the true single organism, preventing artificial inflation of microbial diversity.

NCBI RefSeq / GenBank Accession & Coordinate Provenance

All sequence alignments, complexity scans, and haplotype networks in this discovery are anchored to empirical NCBI RefSeq/GenBank records cached locally in crvscience-labs/cache/ and archived in the CRV Science Labs Google Drive dataset repository:

| Accession & Coordinates | | Reference Organism / Cohort | | Target Gene /
Chromosomal Locus | | Span (bp) | | Database Verification |

| NC_000913.3:4035531-4037072 | | Escherichia coli str. K-12 substr. MG1655 | | rrnA / rrnC / rrnE 16S SSU rRNA | | 1542 | | NCBI RefSeq Verified |

| NC_000913.3:4166659-4168200 | | Escherichia coli str. K-12 substr. MG1655 | | rrnB 16S SSU rRNA Paralog | | 1542 | | NCBI RefSeq Verified |

| NC_000913.3:223771-225312 | | Escherichia coli str. K-12 substr. MG1655 | | rrnH / rrnD 16S SSU rRNA Paralog | | 1542 | | NCBI RefSeq Verified |

| NC_003197.2:4199911-4201452 | | Salmonella enterica serovar Typhimurium LT2 | | rrnA-rrnH 16S SSU rRNA | | 1542 | | NCBI RefSeq Verified |

| NC_000964.3:9809-11362 || Bacillus subtilis subsp. subtilis str. 168 || rrnO /
rrnA-rrnB 16S SSU rRNA || 1554 || NCBI RefSeq Verified |

Generated by: NCBI Entrez E-Utilities ([efetch.fcgi](#)) * CRV Science Labs
PoliteCacheManager ([fetch.rs](#))

References & Primary Sources

* [1] V trovsk T., Baldrian P. (2013). The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. PLOS ONE 8(2):e57923. <https://doi.org/10.1371/journal.pone.0057923>

* [2] Acinas S.G. et al. (2004). Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. Journal of Bacteriology 186(9):2629-2635. <https://doi.org/10.1128/JB.186.9.2629-2635.2004>