DOI: https://doi.org/10.61841/hw0g6q48 Publication URL: https://jarmhs.com/index.php/mhs/article/view/629

# INTRASPECIES CLONAL DIFFERENTIATION OF ESCHERICHIA COLI BY AMPLIFIED RIBOSOMAL DNA RESTRICTION ANALYSIS OF THE 23S RRNA GENE

**[1]Joy Maria Mitchell, [2]Dr. Mohammad Faheem**
*[1,2]Glocal School of Life and Allied Health Sciences*

***Corresponding Author:***

**To Cite This Article:** Mitchell, J. M., & Faheem, M. (2025). INTRASPECIES CLONAL DIFFERENTIATION OF ESCHERICHIA COLI BY AMPLIFIED RIBOSOMAL DNA RESTRICTION ANALYSIS OF THE 23S RRNA GENE. Journal of Advanced Research in Medical and Health Science, 11(8), 33-35. https://doi.org/10.61841/hw0g6q48

## ABSTRACT

*Establishing clonal relationships within Escherichia coli conventionally requires pulsed-field gel electrophoresis, serotyping, or multilocus sequence typing, none of which is inexpensive or rapid. This study tested whether restriction polymorphism within a single conserved core gene can substitute for these methods. Fifty-seven complete E. coli genomes and four non-candidate enterobacterial genomes were mined, 430 copies of the 23S rRNA gene extracted with custom Perl programs, and restriction profiles computed against 199 type II endonucleases grouped into 85 recognition classes. HaeIII emerged as the only enzyme with polymorphic sites that were both conserved in position and variable in occurrence at nucleotides 2161 and 2205, mapping to helices 77 and 78 of domain V. A broad-range primer pair, 23SP2682, was designed to encompass these sites, along with two invariant control sites. The presence or absence of the polymorphic pair defined four restriction types with characteristic fragment profiles on agarose gels. Type I predominated in 32 of 57 genomes (56%), followed by type III (31%), type IV (10%), and type II (2%). Clustering of 71 amplicon sequences reproduced the four groups with minimal overlap, and the types corresponded closely to established phylogroups and pathotypes, with commensals falling in phylogroups A and B1 and pathogens in B2, D, and E. Across 75 field isolates, 93% carried type I profiles.*

**KEYWORDS:** *Escherichia coli; 23S rRNA; ARDRA; clonal typing; phylogroups; restriction polymorphism*

© Author(s) 2025. This work is distributed under the Creative Commons BY- 4.0 license: https://creativecommons.org/licenses/by/4.0/

## 1. INTRODUCTION

Escherichia coli occupies an unusual position in bacteriology: a gut commensal, a laboratory workhorse and an engineered therapeutic chassis, yet also the agent of urinary, enteric and extraintestinal disease. Its population is structured into phylogroups whose assignment has been progressively refined by quadruplex PCR schemes (Clermont et al., 2013), and multilocus sequence analysis has shown that recombination rather than clonal descent alone shapes the relationship between virulence and lineage (Wirth et al., 2006). Comparative genomics has mapped the species onto an open pangenome in which commensal and pathogenic isolates share a modest core surrounded by large accessory repertoires (Rasko et al., 2008), and the phylogeny built on that core has been revised repeatedly as sampling has widened (Chaudhuri and Henderson, 2012).

Resolving clonal relationships in practice remains laborious. Pulsed-field gel electrophoresis, for which interpretive criteria were standardized three decades ago (Tenover et al., 1995), together with plasmid profiling, serotyping, and multilocus sequence typing, requires specialized equipment, extended protocols, or sequencing capacity. A method for reading lineage from one conserved locus by restriction digestion alone would be cheaper, provided that the locus carries variation that is both informative and stable.

Ribosomal operons are attractive candidates — universally present, multi-copy and partly conserved — but divergence between operon copies within one genome is common (Acinas et al., 2004) and has been quantified across hundreds of prokaryotes (Pei et al., 2010). Most studies have used the 16S gene, although the larger 23S gene was shown early on to carry comparable phylogenetic signal (Ludwig and Schleifer, 1994). The present study asks whether restriction polymorphism in 23S rDNA can resolve intraspecies clonal structure in E. coli, and whether the resulting groups correspond to phylogroups and pathotypes assigned by independent methods.

## 2. MATERIALS AND METHODS

Complete genomes of 57 E. coli strains (21 commensal, 1 environmental, 35 pathogenic across 8 pathotypes) and 4 non-candidate enterobacteria were retrieved from GenBank, with accession numbers, genome sizes, GC contents, operon numbers and orientations, sources, pathotypes, and phylogroups tabulated for each. Three Perl programs were written for the analysis: stretch.pl, which extracts a defined coordinate range from a genome file and reports segment size; cutter.pl, which computes restriction profiles; and checkprimer.pl, which verifies primer coverage and fails on a single mismatch. A total of 430 copies of the 23S rRNA gene were extracted, 399 of which were from E. coli genomes.

A panel of 199 type II endonucleases, collapsed into 85 recognition groups, was screened with cutter.pl against reference segments and then across all 430 copies. Candidate sites were located within the 23S secondary structure using Genebee and cross-referenced against published reference structures, retaining only consistently positioned sites. Broad-range primers were designed against the rrlH sequence of E. coli K-12 MG1655 using Primer-BLAST: 23SP2682 (forward 5′-CCGACCTGCACGAATGGCGT-3′, reverse 5′-CAGTTCTCCAGCGCCCACGG-3′), binding helix 71 in domain IV and the helix 93 to 94 transition.

Validation used reference strains representing the K-12, B, C, W, and Crook's lineages, together with 75 field isolates, in equal numbers from clinical, commensal, and environmental sources. The template was prepared by boiling lysis; the reactions used 0.5 U Taq polymerase and 25 cycles at 95 °C, 66 °C, and 72 °C. Amplicons were digested with HaeIII for 1 h at 37 °C and resolved on a 2.5% agarose gel. Seventy-one amplicon sequences were aligned in ClustalW, and the resulting dendrogram was compared with the whole-genome dendrogram for the same 57 strains.

## 3. RESULTS

Enzyme and site selection. Of the 85 recognition groups screened, 43 cut the target, and only 14 produced species-discriminating profiles. Most then failed the conservation screen across all 430 copies, with the site either absent from the candidate species, randomly variable among operon copies within single genomes, or placed in a context that excluded a single-mutation origin. HaeIII was retained because its sites at nucleotides 2161 and 2205 consistently varied between genomes, while their flanking context was invariant. Secondary-structure analysis placed the first at the stem–loop base of helix 77 and the second within the stem of helix 78, both in domain V, a G2163A transition removing the first in a subset of genomes. The polymorphism is not recorded in the Ribosomal RNA Mutation Database.

Definition of restriction types. The 682 bp amplicon included two polymorphic sites, one invariant site upstream, and one downstream, ensuring that complete digestion was always verifiable. Four types follow from the presence or absence of the variable pair: type I lacks both, type IV carries both, types II and III lack one each. On gel, these resolved as 431 bp and 223 bp for type I; 223 bp, 210 bp, and 179 bp for type III; and 223 bp with a 210 bp doublet migrating as a single band for type IV.

Distribution across genomes and isolates. Type I predominated, comprising 32 of 57 genomes (56%), including 17 non-pathogenic and 16 pathogenic strains, followed by type III with 18 genomes (31%), type IV with 6 (10%), and type II with 1 (2%). Seven genomes showed heterogeneity among operon copies, always in a minority pattern, consistent with the intragenomic divergence generally reported for ribosomal operons (Pei et al., 2010). All K-12, C, W, and Crook's strains and their descendants were type I, whereas the B lineage accounted for four of the six type IV genomes. In the field panel, 25 of 25 environmental, 21 of 22 commensal, and 16 of 18 clinical isolates were type I — 93% overall — with type III in two clinical and one commensal isolate, and type II absent.

© Author(s) 2025. This work is distributed under the Creative Commons BY- 4.0 license:
https://creativecommons.org/licenses/by/4.0/

Correspondence with phylogeny. Clustering of 71 amplicon sequences produced four discrete groups with minimal overlap, and the same segregation was recovered from the whole-genome dendrogram, indicating that a pattern read from 682 bp of core sequence is coherent at the genome scale. The types tracked phylogroup closely: A carried type I and, exclusively among B-form strains, type IV, with type I distributed A (n = 15) > B1 (13) > E (7), while type III predominated in B2 and was equally frequent in D. Commensals fell predominantly in A and B1 and pathogens in B2, D and E, matching assignments from independent typing schemes (Clermont et al., 2013). At the pathotype level, laboratory, commensal, enteroaggregative, and enteropathogenic strains were type I; uropathogenic and avian pathogenic strains were type III; and extraintestinal strains were dominated by type IV. Strains at the ancestral root of the dendrogram carried type I profiles, and the first divergence node resolved into maximally separated type I and type IV lineages.

## 4. DISCUSSION AND CONCLUSION

The central finding is that two nucleotide positions in domain V of a ribosomal gene recover a lineage structure that would otherwise require whole-genome or multilocus analysis. The correspondence with phylogroups is close enough to be useful: type I maps onto the commensal-dominated A and B1 groups, type III onto the pathogen-associated B2 and D groups, type IV onto the B lineage. Because the assignment follows from fragment sizes in a single gel lane, the information is obtained in hours and without sequencing.

This does not displace established typing. Pulsed-field gel electrophoresis resolves at a finer grain, and its interpretive criteria are settled (Tenover et al., 1995), while multilocus approaches capture recombination that a single locus cannot see (Wirth et al., 2006). A four-state marker cannot separate strains within a type. The appropriate role is triage: a first-pass assignment that narrows and isolates merit for fuller characterization.

Two features of the data deserve emphasis. First, the intragenomic heterogeneity observed across seven genomes confirms that operon copies within a genome are not interchangeable (Acinas et al., 2004), so profiles should be interpreted as dominant patterns rather than as single genotypes. Second, the dendrogram topology permits an evolutionary reading — two ancestral clonal types with type I dominant, type III as an intermediate arising by acquisition or loss of one site, and type II as a rare configuration — but this rests on one locus and must be tested against the genome-scale phylogenies that have repeatedly revised the species tree (Chaudhuri and Henderson, 2012). It is offered as a hypothesis, not a phylogeny.

The approach is not specific to E. coli or to the 23S gene, though the larger gene offers more positions to work with than the 16S marker at comparable phylogenetic resolution (Ludwig and Schleifer, 1994). Any polymorphism that creates or destroys a recognition sequence within a conserved core gene can be converted into a typing key by the same three-step screen of discrimination, conservation, and structural context.

## REFERENCES

1. Acinas, S.G. et al. (2004) Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. Journal of Bacteriology, 186(9), pp. 2629–2635. doi:10.1128/JB.186.9.2629-2635.2004
2. Chaudhuri, R.R. and Henderson, I.R. (2012) The evolution of the Escherichia coli phylogeny. Infection, Genetics and Evolution, 12(2), pp.214–226. doi:10.1016/j.meegid.2012.01.005
3. Clermont, O. et al. (2013) The Clermont Escherichia coli phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. Environmental Microbiology Reports, 5(1), pp. 58–65. doi:10.1111/1758-2229.12019
4. Ludwig, W. and Schleifer, K.H. (1994) Bacterial phylogeny based on 16S and 23S rRNA sequence analysis. FEMS Microbiology Reviews, 15(2–3), pp. 155–173. doi:10.1111/j.1574-6976.1994.tb00132.x
5. Pei, A.Y. et al. (2010) Diversity of 16S rRNA genes within individual prokaryotic genomes. Applied and Environmental Microbiology, 76(12), pp. 3886–3897. doi:10.1128/AEM.02953-09
6. Rasko, D.A. et al. (2008). The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates. Journal of Bacteriology, 190(20), pp. 6881–6893. doi:10.1128/JB.00619-08
7. Tenover, F.C. et al. (1995) Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. Journal of Clinical Microbiology, 33(9), pp. 2233–2239. doi:10.1128/jcm.33.9.2233-2239.1995
8. Wirth, T. et al. (2006) Sex and virulence in Escherichia coli: an evolutionary perspective. Molecular Microbiology, 60(5), pp. 1136–1151. doi:10.1111/j.1365-2958.2006.05172. x

© Author(s) 2025. This work is distributed under the Creative Commons BY- 4.0 license:
https://creativecommons.org/licenses/by/4.0/