

Summary

Reads for *Escherichia coli* E-2 were submitted to the comprehensive genome analysis service at PATRIC^[1]. Based on the annotation statistics and a comparison to other genomes in PATRIC within this same species, this genome appears to be of Good quality. Details of the analysis, including genes of interest (Specialty Genes), a functional categorization (Subsystems), and a phylogenetic tree (Phylogenetic Analysis) are provided below.

Genome Assembly

Escherichia coli E-2 was assembled using auto^[2]. There were 45 contigs, an estimated genome length of 4,879,867 bp, and an average G+C content of 50.70%. The N50 length, which is defined as the shortest sequence length at 50% of the genome, is 359,891 bp. The L50 count, which is defined as the smallest number of contigs whose length sum produces N50, is 5 (Table 1).

Table 1. Assembly Details	
Contigs	45
GC Content	50.70
Plasmids	0
Contig L50	5
Genome Length	4,879,867 bp
Contig N50	359,891
Chromosomes	0
Job ID	assembly_873911
Job Started	February 9th 2025, 2:17:19pm
Job Completed	February 9th 2025, 3:14:24pm
Total Time	57 minutes and 5 seconds
Selected Recipe	auto

Genome Annotation

The *Escherichia coli* E-2 genome was annotated using RAST tool kit (RASTtk)^[3] and assigned a unique genome identifier of 562.162465. This genome is in the superkingdom Bacteria and was annotated using genetic code 11. The taxonomy of this genome is:

cellular organisms > Bacteria > Pseudomonadati > Pseudomonadota > Gammaproteobacteria > Enterobacterales > Enterobacteriaceae > *Escherichia* > *Escherichia coli*

This genome has 4,749 protein coding sequences (CDS), 71 transfer RNA (tRNA) genes, and 3 ribosomal RNA (rRNA) genes. The annotated features are summarized in Table 2.

Table 2. Annotated Genome Features	
CDS	4,749
tRNA	71
rRNA	3
Partial CDS	0
Miscellaneous RNA	0
Repeat Regions	0
Job ID	annotation_873911
Job Started	February 9th 2025, 3:14:24pm
Job Completed	February 9th 2025, 3:21:24pm
Total Time	7 minutes and 0 seconds

Activate Windows
Go to Settings to activate Windows.

The annotation included 434 hypothetical proteins and 4,315 proteins with functional assignments (Table 3). The proteins with functional assignments included 1,285 proteins with Enzyme Commission (EC) numbers^[4], 1,072 with Gene Ontology (GO) assignments^[5], and 904 proteins that were mapped to KEGG pathways^[6]. PATRIC annotation includes two types of protein families^[7], and this genome has 4,611 proteins that belong to the genus-specific protein families (PLFams) for , and 4,663 proteins that belong to the cross-genus protein families (PGFams).

Table 3. Protein Features	
Hypothetical proteins	434
Proteins with functional assignments	4,315
Proteins with EC number assignments	1,285
Proteins with GO assignments	1,072
Proteins with Pathway assignments	904
Proteins with PATRIC genus-specific family (PLfam) assignments	4,611
Proteins with PATRIC cross-genus family (PGfam) assignments	4,663

A circular graphical display of the distribution of the genome annotations is provided (Figure 1). This includes, from outer to inner rings, the contigs, CDS on the forward strand, CDS on the reverse strand, RNA genes, CDS with homology to known antimicrobial resistance genes, CDS with homology to know virulence factors, GC content and GC skew. The colors of the CDS on the forward and reverse strand indicate the subsystem that these genes belong to (see Subsystems below).

Figure 1

