Report Date: 02/09/2025
Organism: Escherichia coli E-1

Summary

Reads for Escherichia coli E-1 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-1 was assembled using auto[2]. There were 243 contigs, an estimated genome length of 5,319,590 bp, and an average G+C content of 50.28%. The N50 length, which is defined as the shortest sequence length at 50% of the genome, is 146,357 bp. The L50 count, which is defined as the smallest number of contigs whose length sum produces N50, is 12 (**Table 1**).

Table 1. Assembly Details

Contigs	243
GC Content	50.28
Plasmids	0
Contig L50	12
Genome Length	5,319,590 bp
Contig N50	146,357
Chromosomes	0
Job ID	assembly_1237610
Job Started	February 9th 2025, 2:51:59pm
Job Completed	February 9th 2025, 4:44:10pm
Total Time	1h52m11s
Selected Recipe	auto

Genome Annotation

The Escherichia coli E-1 genome was annotated using RAST tool kit (RASTtk)[3] and assigned a unique genome identifier of 562.162467. This genome is in the superkingdom Bacteria and was annotated using genetic code 11. The taxonomy of this genome is:

Genome Annotation

The Escherichia coli E-1 genome was annotated using RAST tool kit (RASTtk)[3] and assigned a unique genome identifier of 562.162467. 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 5,518 protein coding sequences (CDS), 90 transfer RNA (tRNA) genes, and 4 ribosomal RNA (rRNA) genes. The annotated features are summarized in **Table 2**.

Table 2. Annotated Genome Features

CDS	5,518
tRNA	90
rRNA	4
Repeat Regions	4
Partial CDS	0
Miscellaneous RNA	0
Job ID	annotation_1237610
Job Started	February 9th 2025, 4:44:10pm
Job Completed	February 9th 2025, 4:50:52pm
Total Time	6 minutes and 42 seconds

The annotation included 645 hypothetical proteins and 4,873 proteins with functional assignments (**Table 3**). The proteins with functional assignments included 1,300 proteins with Enzyme Commission (EC) numbers[4], 1,063 with Gene Ontology (GO) assignments[5], and 884 proteins that were mapped to KEGG pathways[6]. PATRIC annotation includes two types of protein families[7], and this genome has 5,308 proteins that belong to the genus-specific protein families (PLFams) for , and 5,387 proteins that belong to the cross-genus protein families (PGFams).

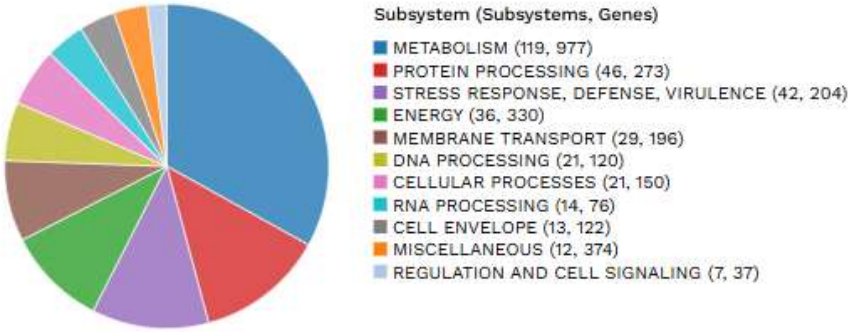
Table 3. Protein Features

Hypothetical proteins	645
Proteins with functional assignments	4,873
Proteins with EC number assignments	1,300
Proteins with GO assignments	1,063
Proteins with Pathway assignments	884
Proteins with PATRIC genus-specific family (PLfam) assignments	5,308
Proteins with PATRIC cross-genus family (PGfam) assignments	5,387

Subsystem Analysis

A subsystem is a set of proteins that together implement a specific biological process or structural complex[8] and PATRIC annotation includes an analysis of the subsystems unique to each genome. An overview of the subsystems for this genome is provided in **Figure 2**.

Figure 2



Specialty Genes

Many of the genes annotated in have homology to known transporters[9], virulence factors[10][11], drug targets[12][13], and antibiotic resistance genes[14]. The number of genes and the specific source database where homology was found is provided (**Table 4**).

Table 4. Specialty Genes

	Source	Genes
	Victors	2
Antibiotic Resistance	CARD	78
Antibiotic Resistance	NDARO	1
Antibiotic Resistance	PATRIC	65
Drug Target	DrugBank	385
Drug Target	TTD	62
Transporter	TCDB	898
Virulence Factor	PATRIC_VF	290
Virulence Factor	VFDB	141
Virulence Factor	Victors	302

Antimicrobial Resistance Genes

The Genome Annotation Service in PATRIC uses k-mer-based AMR genes detection method, which utilizes PATRIC's curated collection of representative AMR gene sequence variants[1] and assigns to each AMR gene functional annotation, broad mechanism of antibiotic resistance, drug class and, in some cases, specific antibiotic it confers resistance to. Please note, that the presence of AMR-related genes (even full length) in a given genome does not directly imply antibiotic resistant phenotype. It is important to consider specific AMR mechanisms and especially the absence/presence of SNP mutations conveying resistance. A summary of the AMR genes annotated in this genome and corresponding AMR mechanism is provided in **Table 5**.

Table 5. Antimicrobial Resistance Genes

AMR Mechanism	Genes
Antibiotic activation enzyme	KatG
Antibiotic inactivation enzyme	BlaEC family
Antibiotic resistance gene cluster,cassette,or operon	MarA, MarB, MarR
Antibiotic target in susceptible species	Alr, Ddl, dxr, EF-G, EF-Tu, folA, Dfr, folP, gyrA, gyrB, inhA, fabI, Iso-tRNA, kasA, MurA, rho, rpoB, rpoC, S10p, S12p
Antibiotic target protection protein	BcrC
Efflux pump conferring antibiotic resistance	AcrAB-TolC, AcrAD-TolC, AcrEF-TolC , AcrZ, EmrAB-TolC, EmrD, EmrE, EmrKY-TolC, MacA, MacB, MdfA/Cmr, MdtABC-TolC, MdtEF-TolC, MdtL, MdtM, SugE, TolC/OpmH
Gene conferring resistance via absence	gidB
Protein altering cell wall charge conferring antibiotic resistance	GdpD, PgsA
Regulator modulating expression of antibiotic resistance genes	AcrAB-TolC, EmrAB-TolC, GadE, H-NS, OxyR

Phylogenetic Analysis

The National Center for Biotechnology Information (NCBI) staff manually select and categorize reference and representative genomes, which they consider to be of high quality and importance to the research community. PATRIC provides the reference and representative genomes, and includes them in the phylogenetic analysis that is part of the Comprehensive Genome Analysis report. The closest reference and representative genomes to were identified by Mash/MinHash[16]. PATRIC global protein families (PGFams)[7] were selected from these genomes to determine the phylogenetic placement of this genome. The protein sequences from these families were aligned with MUSCLE[18], and the nucleotides for each of those sequences were mapped to the protein alignment. The joint set of amino acid and nucleotide alignments were concatenated into a data matrix, and RaxML[19] was used to analyze this matrix, with fast bootstrapping[20] was used to generate the support values in the tree (**Figure 3**).