

Resource Summary Report

Generated by [NIF](#) on Sep 3, 2026

GenProtEC

RRID:SCR_002804

Type: Tool

Proper Citation

GenProtEC (RRID:SCR_002804)

Resource Information

URL: <http://genprotec.mbl.edu>

Proper Citation: GenProtEC (RRID:SCR_002804)

Description: GenProtEC is dedicated to the functions encoded by the Escherichia coli K-12 (strain MG1655) genome defined in the GenBank Accession No. NC_000913.2 deposit. All the data presented in GenProtEC is made easily accessible to the users through downloadable flat files in text format. GenProtEC presents information on the functions of E.coli K-12 MG1655 gene products from several points of view. E.coli proteins as single modules have been grouped in sequence similarity. Using the power of group membership of proteins of similar function, open reading frames within any group can be assigned the general function. In addition, the presence of domains of known function within E.coli proteins has been determined. Domain content permits annotation of some functional information to otherwise totally unknown sequences. The rich classification of cellular roles, MultiFun, has been applied, underlining the fact that many gene products have more than one cellular role. Our annotation work includes multiple types of information: 1. Sequence similarity to orthologues as defined by Darwin (start and end of aligned region, identity, and PAM distance). 2. Resolution of fused proteins into modular units with independent functions. 3. Identification of sequence similar protein groups within E. coli that are clustered by transitive relationships. The sequence similarity is limited to PAM 200 and an alignment of at least 83 amino acids. 4. Updated literature references. 5. Classification of gene products by their gene type and by their cellular role(s). The MultiFun classification system for cellular roles is used to assign gene products to one or more roles. MultiFun has been converted to Gene Ontology terms. 6. Families of proteins related by structure and biochemical reaction mechanisms (work in progress). 7. SCOP superfamily identification and location (e.g. binding site domains) for E. coli proteins., THIS RESOURCE IS NO LONGER IN SERVICE. Documented on January 15,2026.

Abbreviations: GenProtEC

Synonyms: GenProtEC - E. coli genome and proteome database, GenProtEC- E. coli Genome and Proteome Database

Resource Type: data or information resource, database

Defining Citation: [PMID:14681418](#)

Keywords: e. coli, escherichia coli, escherichia coli k-12, bacterial genome

Availability: THIS RESOURCE IS NO LONGER IN SERVICE

Resource Name: GenProtEC

Resource ID: SCR_002804

Alternate IDs: nif-0000-02904

Record Creation Time: 20220129T080215+0000

Record Last Update: 20260829T182919+0000

Ratings and Alerts

No rating or validation information has been found for GenProtEC.

No alerts have been found for GenProtEC.

Data and Source Information

Source: [SciCrunch Registry](#)

Usage and Citation Metrics

We found 18 mentions in open access literature.

Listed below are recent publications. The full list is available at [NIF](#) .

Santos-Zavaleta A, et al. (2018) A unified resource for transcriptional regulation in Escherichia coli K-12 incorporating high-throughput-generated binding data into RegulonDB version 10.0. BMC biology, 16(1), 91.

Pawar KI, et al. (2017) Role of D-aminoacyl-tRNA deacylase beyond chiral proofreading as a cellular defense against glycine mischarging by AlaRS. eLife, 6.

Murakami Y, et al. (2015) Global coordination in adaptation to gene rewiring. Nucleic acids research, 43(2), 1304.

Cavalcanti JH, et al. (2014) Evolution and functional implications of the tricarboxylic acid cycle as revealed by phylogenetic analysis. Genome biology and evolution, 6(10), 2830.

Marisch K, et al. (2013) A comparative analysis of industrial Escherichia coli K-12 and B strains in high-glucose batch cultivations on process-, transcriptome- and proteome level. PloS one, 8(8), e70516.

Edwards AL, et al. (2013) Transient growth arrest in *Escherichia coli* induced by chromosome condensation. *PloS one*, 8(12), e84027.

Suez J, et al. (2013) Virulence gene profiling and pathogenicity characterization of non-typhoidal *Salmonella* accounted for invasive disease in humans. *PloS one*, 8(3), e58449.

Wernegreen JJ, et al. (2011) Reduced selective constraint in endosymbionts: elevation in radical amino acid replacements occurs genome-wide. *PloS one*, 6(12), e28905.

Fan J, et al. (2011) An efficient strategy for high throughput screening of recombinant integral membrane protein expression and stability. *Protein expression and purification*, 78(1), 6.

Warnecke T, et al. (2010) GroEL dependency affects codon usage--support for a critical role of misfolding in gene evolution. *Molecular systems biology*, 6, 340.

Scharl T, et al. (2009) Exploratory and inferential analysis of gene cluster neighborhood graphs. *BMC bioinformatics*, 10, 288.

Freyre-González JA, et al. (2008) Functional architecture of *Escherichia coli*: new insights provided by a natural decomposition approach. *Genome biology*, 9(10), R154.

Liang H, et al. (2008) Protein under-wrapping causes dosage sensitivity and decreases gene duplicability. *PLoS genetics*, 4(1), e11.

Malik A, et al. (2007) Databases and QSAR for cancer research. *Cancer informatics*, 2, 99.

Riley M, et al. (2006) *Escherichia coli* K-12: a cooperatively developed annotation snapshot--2005. *Nucleic acids research*, 34(1), 1.

Galperin MY, et al. (2005) The Molecular Biology Database Collection: 2005 update. *Nucleic acids research*, 33(Database issue), D5.

Voit EO, et al. (2003) Extending knowledge of *Escherichia coli* metabolism by modeling and experiment. *Genome biology*, 4(11), 235.

Serres MH, et al. (2001) A functional update of the *Escherichia coli* K-12 genome. *Genome biology*, 2(9), RESEARCH0035.