

# BSgenome.Ecoli.NCBI.20080805

June 3, 2026

---

BSgenome.Ecoli.NCBI.20080805

*Escherichia coli* full genomes

---

## Description

Escherichia coli full genomes for several strains as provided by NCBI on 2008/08/05 and stored in Biostrings objects.

## Details

The genome of Escherichia coli is made of a single circular DNA sequence. The full genomes for the following strains are present in this package (one sequence per strain):

### Escherichia coli 536

Taxonomy ID: 362663

Other names: "Escherichia coli strain 536", "Escherichia coli str. 536"

Refseq: NC\_008253

Length: 4,938,920 nt

Seq.Status: Completed

Sequencing center: University of Goettingen

Completed: 2006/07/24

File: ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Escherichia\_coli\_536/NC\_008253.fna

### Escherichia coli APEC O1

Taxonomy ID: 405955

Other names: "Escherichia coli strain APEC O1", "Escherichia coli str. APEC O1"

Refseq: NC\_008563

Length: 5,082,025 nt

Seq.Status: Completed

Sequencing center: Iowa State University

Completed: 2006/11/08

File: ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Escherichia\_coli\_APEC\_O1/NC\_008563.fna

### Escherichia coli ATCC 8739

Taxonomy ID: 481805  
Other names: "Escherichia coli C (ATCC 8739)", "Escherichia coli C str. ATCC 8739", "Escherichia coli C"  
Refseq: NC\_010468  
Length: 4,746,218 nt  
Seq.Status: Completed  
Sequencing center: US DOE Joint Genome Institute (JGI-PGF)  
Completed: 2008/03/17  
File: ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Escherichia\_coli\_C\_ATCC\_8739/NC\_010468.fna

#### Escherichia coli CFT073

Taxonomy ID: 199310  
Other names: "Escherichia coli strain CFT073", "Escherichia coli str. CFT073"  
Refseq: NC\_004431  
Length: 5,231,428 nt  
Seq.Status: Completed  
Sequencing center: Univ. Wisconsin  
Completed: 2002/12/09  
File: ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Escherichia\_coli\_CFT073/NC\_004431.fna

#### Escherichia coli E24377A

Taxonomy ID: 331111  
Refseq: NC\_009801  
Length: 4,979,619 nt  
Seq.Status: Completed  
Sequencing center: TIGR  
Completed: 2007/09/13  
File: ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Escherichia\_coli\_E24377A/NC\_009801.fna

#### Escherichia coli HS

Taxonomy ID: 331112  
Refseq: NC\_009800  
Length: 4,643,538 nt  
Seq.Status: Completed  
Sequencing center: TIGR  
Completed: 2007/09/13  
File: ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Escherichia\_coli\_HS/NC\_009800.fna

#### Escherichia coli O157:H7 EDL933

Taxonomy ID: 155864  
Refseq: NC\_002655  
Length: 5,528,445 nt  
Seq.Status: Completed  
Sequencing center: Univ. Wisconsin  
Completed: 2001/09/27  
File: ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Escherichia\_coli\_O157H7\_EDL933/NC\_002655.fna

#### Escherichia coli O157:H7 str. Sakai

Taxonomy ID: 386585  
Other names: "Escherichia coli O157:H7 strain Sakai"  
Refseq: NC\_002695  
Length: 5,498,450 nt  
Seq.Status: Completed  
Sequencing center: GIRC

Completed: 2001/10/02

File: ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Escherichia\_coli\_0157H7/NC\_002695.fna

Escherichia coli SMS-3-5

Taxonomy ID: 439855

Other names: "Escherichia coli strain SMS-3-5", "Escherichia coli str. SMS-3-5"

Refseq: NC\_010498

Length: 5,068,389 nt

Seq.Status: Completed

Sequencing center: TIGR

Completed: 2008/03/24

File: ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Escherichia\_coli\_SMS\_3\_5/NC\_010498.fna

Escherichia coli UTI89

Taxonomy ID: 364106

Other names: "Escherichia coli strain UTI89", "Escherichia coli str. UTI89"

Refseq: NC\_007946

Length: 5,065,741 nt

Seq.Status: Completed

Sequencing center: Washington University (WashU)

Completed: 2006/04/07

File: ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Escherichia\_coli\_UTI89/NC\_007946.fna

Escherichia coli str. K12 substr. DH10B

Taxonomy ID: 316385

Other names: "Escherichia coli DH10B", "Escherichia coli strain K12 substrain DH10B"

Refseq: NC\_010473

Length: 4,686,137 nt

Seq.Status: Completed

Sequencing center: University of Wisconsin-Madison

Completed: 2008/03/17

File: ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Escherichia\_coli\_K\_12\_substr\_\_DH10B/NC\_010473.fna

Escherichia coli str. K12 substr. MG1655

Taxonomy ID: 511145

Other names: "Escherichia coli MG1655", "Escherichia coli strain MG1655", "Escherichia coli str.

Refseq: NC\_000913

Length: 4,639,675 nt

Seq.Status: Completed

Sequencing center: Univ. Wisconsin

Completed: 2001/10/15

File: ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Escherichia\_coli\_K12\_substr\_\_MG1655/NC\_000913.fna

Escherichia coli str. K12 substr. W3110

Taxonomy ID: 316407

Other names: "Escherichia coli W3110", "Escherichia coli strain W3110", "Escherichia coli str. W

Refseq: AC\_000091

Length: 4,646,332 nt

Seq.Status: Completed

Sequencing center: Nara Institute of Science and Technology

Completed: 2006/03/01

File: ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Escherichia\_coli\_W3110/AC\_000091.fna

**Note**

This BSgenome data package was made from the following source data files:

```
Escherichia_coli_536/NC_008253.fna
Escherichia_coli_APEC_01/NC_008563.fna
Escherichia_coli_C_ATCC_8739/NC_010468.fna
Escherichia_coli_CFT073/NC_004431.fna
Escherichia_coli_E24377A/NC_009801.fna
Escherichia_coli_HS/NC_009800.fna
Escherichia_coli_0157H7_EDL933/NC_002655.fna
Escherichia_coli_0157H7/NC_002695.fna
Escherichia_coli_SMS_3_5/NC_010498.fna
Escherichia_coli_UTI89/NC_007946.fna
Escherichia_coli_K_12_substr__DH10B/NC_010473.fna
Escherichia_coli_K12_substr__MG1655/NC_000913.fna
Escherichia_coli_W3110/AC_000091.fna
from ftp://ftp.ncbi.nih.gov/genomes/Bacteria/ (downloaded on 2008/08/05)
```

See [?BSgenomeForge](#) and the BSgenomeForge vignette (`vignette("BSgenomeForge")`) in the **BSgenome** software package for how to make a BSgenome data package.

**Author(s)**

The Bioconductor Dev Team

**See Also**

- [BSgenome](#) objects and the `available.genomes` function in the **BSgenome** software package.
- [DNAString](#) objects in the **Biostrings** package.
- The BSgenomeForge vignette (`vignette("BSgenomeForge")`) in the **BSgenome** software package for how to make a BSgenome data package.

**Examples**

```
BSgenome.Ecoli.NCBI.20080805
genome <- BSgenome.Ecoli.NCBI.20080805
seqlengths(genome)
## The names of the sequences are the Refseq numbers!
genome$NC_008253 # same as genome[["NC_008253"]]

## -----
## Genome-wide motif searching
## -----
## See the GenomeSearching vignette in the BSgenome software
## package for some examples of genome-wide motif searching using
## Biostrings and the BSgenome data packages:
if (interactive())
  vignette("GenomeSearching", package="BSgenome")
```

# Index

**\* data**

BSgenome.Ecoli.NCBI.20080805, [1](#)

**\* package**

BSgenome.Ecoli.NCBI.20080805, [1](#)

available.genomes, [4](#)

BSgenome, [4](#)

BSgenome.Ecoli.NCBI.20080805, [1](#)

BSgenome.Ecoli.NCBI.20080805-package  
(BSgenome.Ecoli.NCBI.20080805),  
[1](#)

BSgenomeForge, [4](#)

DNAStrng, [4](#)

Ecoli (BSgenome.Ecoli.NCBI.20080805), [1](#)