

# NCBI ftp server for microbial genomes

# ftp – file transfer protocol

Made to transfer files from one computer to  
another


modern web browsers support protocol



**http**: hypertext transfer protocol

WWW build by hyperlinks

# http://www.ncbi.nlm.nih.gov/Ftp/

 **NCBI FTP site**

PubMed Entrez BLAST OMIM Books TaxBrowser Structure

Search  for

**NCBI**  
[SITE MAP](#)  
Guide to NCBI resources

**About NCBI**  
The science behind our resources. An introduction for researchers, educators and the public.

**GenBank**  
sequence submission support and software

**Molecular databases**  
sequences, structures and taxonomy

**Literature databases**  
PubMed and OMIM

**Genomic Biology**  
the human genome, whole genomes and related resources

**Tools**  
for data mining

**Research at NCBI**  
people, projects and seminars

**Major resources available by ftp ([ftp.ncbi.nih.gov](http://ftp.ncbi.nih.gov)):**

▶ [BLAST Basic Local Alignment Search Tool](#)

Download the BLAST database and stand-alone sequence comparison software.

▶ [CDD Data](#)

Download data from the Conserved Domain Database.

▶ [CD-Tree](#)

Download the protein domain hierarchy viewer and editor.

▶ [Cn3D](#)

Download the stand-alone software for viewing 3-dimensional structures.

▶ [Data Repository](#)

Download collections of contributed molecular biology data.

▶ [dbGaP](#)

Download open access Genotype and Phenotype data.

▶ [GenBank](#)

Download the full release database, daily updates, or WGS files.

Note: there is a mirror site for GenBank files at Indiana University ([bio-mirror.net/biomirror/genbank](http://bio-mirror.net/biomirror/genbank)).

▶ [Gene](#)

Download gene-based information from completely sequenced organisms.

▶ [Genome Assembly/Annotation Projects](#)

Download complete genomes/chromosomes, contigs and reference sequence mRNAs and proteins.

▶ [MMDB](#)

Download NCBI's structure database.

people, projects and seminars

**Software engineering**  
tools, R&D and databases

**Education**  
teaching resources and on-line tutorials

**Contact information**  
how to reach us

▶ [MMDB](#)

Download NCBI's structure database.

▶ [NCBI Toolbox](#)

Download the NCBI software tools for building bioinformatics resources.

▶ [PubChem](#)

Download the database of chemical structures of small organic molecules and information on their biological activities.

▶ [RefSeq](#)

Download the curated RefSeq full release or daily updates.

▶ [Sequin](#)

Download the stand-alone GenBank sequence submission software.

▶ [dbSNP](#)

Download the database of short genetic variations.

▶ [Taxonomy](#)

Download data files from the Taxonomy database.

▶ [UniGene](#)

Download data files from the UniGene datasets of non-redundant gene-oriented clusters.

▶ [UniSTS](#)


Download data files from the UniSTS resource of sequence tagged sites.

Information on [how to uncompress](#) \*.gz and \*.Z files is also available.

[Disclaimer](#) [Privacy statement](#)

Revised: January 13, 2009.

# http://www.ncbi.nlm.nih.gov/Ftp/

 **NCBI FTP site**

PubMed Entrez BLAST OMIM Books TaxBrowser Structure

Search All Databases for Go

**NCBI**  
SITE MAP  
Guide to NCBI resources

About NCBI  
The science behind our resources. An introduction for researchers, educators and the public.

GenBank  
sequence submission support and software

Molecular databases  
sequences, structures and taxonomy

Literature databases  
PubMed and OMIM

Genomic Biology  
the human genome, whole genomes and related resources

Tools  
for data mining

Research at NCBI  
people, projects and seminars

**Major resources available by ftp ([ftp.ncbi.nih.gov](http://ftp.ncbi.nih.gov)):**

[BLAST Basic Local Alignment Search Tool](#)  
Download the BLAST database and stand-alone sequence comparison software.

[CDD Data](#)  
Download data from the Conserved Domain Database.

[CD-Tree](#)  
Download the protein domain hierarchy viewer and editor.

[Cn3D](#)  
Download the stand-alone software for viewing 3-dimensional structures.

[Data Repository](#)  
Download collections of contributed molecular biology data.

[dbGaP](#)  
Download open access Genotype and Phenotype data.

[GenBank](#) 2  
Download the full release database, daily updates, or WGS files.  
Note: there is a mirror site for GenBank files at Indiana University ([bio-mirror.net/biomirror/genbank](http://bio-mirror.net/biomirror/genbank)).

[Gene](#)  
Download gene-based information from completely sequenced organisms.

[Genome Assembly/Annotation Projects](#) 1  
Download complete genomes/chromosomes, contigs and reference sequence mRNAs and proteins.

[MMDB](#)  
Download NCBI's structure database.

people, projects and seminars

Software engineering  
tools, R&D and databases

Education  
teaching resources and on-line tutorials

Contact information  
how to reach us

[MMDB](#)  
Download NCBI's structure database.

[NCBI Toolbox](#)  
Download the NCBI software tools for building bioinformatics resources.

[PubChem](#)  
Download the database of chemical structures of small organic molecules and information on their biological activities.

[RefSeq](#)  
Download the curated RefSeq full release or daily updates.

[Sequin](#)  
Download the stand-alone GenBank sequence submission software.

[dbSNP](#)  
Download the database of short genetic variations.

[Taxonomy](#)  
Download data files from the Taxonomy database.

[UniGene](#)  
Download data files from the UniGene datasets of non-redundant gene-oriented clusters.


[UniSTS](#)  
Download data files from the UniSTS resource of sequence tagged sites.

Information on [how to uncompress](#) \*.gz and \*.Z files is also available.

[Disclaimer](#) [Privacy statement](#)

Revised: January 13, 2009.

# http://www.ncbi.nlm.nih.gov/Ftp/

 **NCBI FTP site**

PubMed Entrez BLAST OMIM Books TaxBrowser Structure

Search All Databases for  Go

**NCBI**  
SITE MAP  
Guide to NCBI resources

About NCBI  
The science behind our resources. An introduction for researchers, educators and the public.

GenBank  
sequence submission support and software

Molecular databases  
sequences, structures and taxonomy

Literature databases  
PubMed and OMIM

Genomic Biology  
the human genome, whole genomes and related resources

Tools  
for data mining

Research at NCBI  
people, projects and seminars

**Major resources available by ftp ([ftp.ncbi.nih.gov](http://ftp.ncbi.nih.gov)):**

[BLAST Basic Local Alignment Search Tool](#)  
Download the BLAST database and stand-alone sequence comparison software.

[CDD Data](#)  
Download data from the Conserved Domain Database.

[CD-Tree](#)  
Download the protein domain hierarchy viewer and editor.

[Cn3D](#)  
Download the stand-alone software for viewing 3-dimensional structures.

[Data Repository](#)  
Download collections of contributed molecular biology data.

[dbGaP](#)  
Download open access Genotype and Phenotype data.

[GenBank](#) 2  
Download the full release database, daily updates, or WGS files.  
Note: there is a mirror site for GenBank files at Indiana University ([bio-mirror.net/biomirror/genbank](http://bio-mirror.net/biomirror/genbank)).

[Gene](#)  
Download gene-based information from completely sequenced organisms.

[Genome Assembly/Annotation Projects](#) 1  
Download complete genomes/chromosomes, contigs and reference sequence mRNAs and proteins.

[MMDB](#)  
Download NCBI's structure database.

people, projects and seminars

Software engineering  
tools, R&D and databases

Education  
teaching resources and on-line tutorials

Contact information  
how to reach us

[MMDB](#)  
Download NCBI's structure database.

[NCBI Toolbox](#)  
Download the NCBI software tools for building bioinformatics resources.

[PubChem](#)  
Download the database of chemical structures of small organic molecules and information on their biological activities.

[RefSeq](#)  
Download the curated RefSeq full release or daily updates.

[Sequin](#)  
Download the stand-alone GenBank sequence submission software.

[dbSNP](#)  
Download the database of short genetic variations.

[Taxonomy](#)  
Download data files from the Taxonomy database.

[UniGene](#)  
Download data files from the UniGene datasets of non-redundant gene-oriented clusters.

[UniSTS](#)  
Download data files from the UniSTS resource of sequence tagged sites.

Information on [how to uncompress](#) \*.gz and \*.Z files is also available.

[Disclaimer](#) [Privacy statement](#)

Revised: January 13, 2009.




























# Genome Assembly/Annotation Projects

## RefSeq



Index of <ftp://ftp.ncbi.nih.gov/genomes/>

[^ Up to higher level directory](#)

Name	Size	Last Modified
 ASSEMBLY_REPORTS		18.10.2012 02:51:00
 Acyrthosiphon_pisum		15.03.2012 00:00:00
 Aedes_aegypti		21.09.2010 00:00:00
 Ailuropoda_melanoleuca		06.09.2012 20:24:00
 Amphimedon_queenslandica		15.03.2012 00:00:00
 Anolis_carolinensis		15.03.2012 00:00:00
 Anopheles_gambiae		14.01.2008 00:00:00
 Apis_florea		13.04.2012 00:00:00
 Apis_mellifera		14.06.2011 00:00:00
 Arabidopsis_lyrata		11.05.2011 00:00:00
 Arabidopsis_thaliana		30.09.2008 00:00:00
 BACENDS		13.05.2010 00:00:00
 Bacteria		18.10.2012 10:42:00
 Bacteria_DRAFT		10.10.2012 16:14:00
 Bombus_impatiens		15.03.2012 00:00:00
 Bombus_terrestris		15.03.2012 00:00:00
 Bos_taurus		07.03.2012 00:00:00
 Brachypodium_distachyon		15.03.2012 00:00:00
 Branchiostoma_floridae		26.01.2010 00:00:00
 Brugia_malayi		26.09.2007 00:00:00
 CLONEEND		21.12.2006 00:00:00
 CLUSTERS		30.11.2010 00:00:00
 Caenorhabditis_elegans		10.10.2008 00:00:00
 Callithrix_jacchus		12.07.2012 21:11:00
 <i>Canis_familiaris</i>		10.02.2012 00:00:00
 Canis_lupus_familiaris		15.03.2012 00:00:00
 Cavia_porcellus		15.03.2012 00:00:00
 Chloroplasts		01.06.2011 00:00:00

# FileZilla



More comfort with special ftp clients,  
like FileZilla (GPL)

<http://filezilla-project.org/>



[Home](#)

## FileZilla

[Features](#)  
[Screenshots](#)  
[Download](#)  
[Documentation](#)

## FileZilla Server

[Download](#)

## General

[Contact](#)  
[Forum](#)  
[Project page](#)  
[Wiki](#)

## Development

[Source code](#)  
[Nightly builds](#)  
[Translations](#)  
[Version history](#)  
[Changelog](#)  
[Issue tracker](#)

## Other projects

[Octochess](#)



Sponsors:



## Overview

Welcome to the homepage of FileZilla, the free FTP solution. Both a client and a server are available. FileZilla is open source software distributed free of charge under the terms of the GNU General Public License

Support is available through our [forums](#), the [wiki](#) and the [bug and feature request trackers](#).

In addition, you will find documentation on how to compile FileZilla and nightly builds for multiple platforms in the development section.

### Quick download links

**Download  
FileZilla Client**

All platforms

**Download  
FileZilla Server**

Windows only

Pick the client if you want to transfer files. Get the server if you want to make files available for others.

### News

#### 2012-02-26 - FileZilla Server 0.9.41 released

##### Bugfixes and minor changes:

- Fix parsing of IP address filters ending with :0 or equivalent substringss.
- Allow speed limits larger than 64 MiB/s.
- Show more verbose error messages if transfer connection cannot be established.

#### 2012-01-08 - FileZilla Client 3.5.3 released

##### New features:

- Ctrl+Shift+O can be used to toggle between modification time and filesize based directory comparison

##### Bugfixes and minor changes:

# ftp://ftp.ncbi.nih.gov/genomes/



Host: ftp.ncbi.nih.gov Username: anonymou Password: Port: Quickconnect

Command: PASV  
Response: 227 Entering Passive Mode (130,14,250,11,196,215).  
Command: MLSD  
Response: 150 Opening ASCII mode data connection for MLSD  
Response: 226 Transfer complete  
Status: Directory listing successful

Local site: /home/joe/bioperl/MLST/iz\_panel/

- MLST
  - AT\_run
  - achtman\_backup
  - backup
  - fasta\_backups
  - iz\_panel
  - iz\_panel\_MLST\_analysis
  - test
- Perl

Filename	Filesize	Filetype	Last modified
..			
Ecoli_042_uid161985.fasta	5,5 MB	fasta-file	04.10.2012 16:...
Ecoli_101_1_uid54363_draft_con.fasta	5,1 MB	fasta-file	04.10.2012 16:...
Ecoli_1180_uid40289_draft_con.fasta	5,7 MB	fasta-file	04.10.2012 16:...
Ecoli_1357_uid40291_draft_con.fasta	5,4 MB	fasta-file	04.10.2012 16:...
Ecoli_1827_70_uid60615_draft_con.fasta	4,9 MB	fasta-file	04.10.2012 16:...
Ecoli_2362_75_uid60613_draft_con.fasta	5,3 MB	fasta-file	04.10.2012 16:...
Ecoli_3431_uid40265_draft_con.fasta	5,4 MB	fasta-file	04.10.2012 16:...
Ecoli_53638_uid54321_draft_con.fasta	5,2 MB	fasta-file	04.10.2012 16:...
Ecoli_536_uid58531.fasta	5,1 MB	fasta-file	04.10.2012 16:...
Ecoli_55989_uid59383.fasta	5,3 MB	fasta-file	04.10.2012 16:...
Ecoli_83972_uid55485_draft_con.fasta	5,2 MB	fasta-file	04.10.2012 16:...
Ecoli_AA86_uid65321_draft_con.fasta	5,1 MB	fasta-file	04.10.2012 16:...
Ecoli_ABU_83972_uid161975.fasta	5,3 MB	fasta-file	04.10.2012 16:...
Ecoli_ATCC_8739_uid58783.fasta	4,9 MB	fasta-file	04.10.2012 16:...
Ecoli_B088_uid47003_draft_scaf.fasta	5,1 MB	fasta-file	04.10.2012 16:...
Ecoli_B171_uid54319_draft_con.fasta	5,6 MB	fasta-file	04.10.2012 16:...
Ecoli_B185_uid47001_draft_scaf.fasta	5,3 MB	fasta-file	04.10.2012 16:...
Ecoli_B354_uid46999_draft_scaf.fasta	5,0 MB	fasta-file	04.10.2012 16:...

Selected 1 file. Total size: 5,5 MB

Remote site: /genomes

- entrez
- epigenomics
- fa2htgs
- genbank
- gene
- genomes
- geo
- hapmap
- mmdb
- ncbi-asn1

Filename	Filesize	Filetype	Last modified	Permission	Owner/Gro
..					
ASSEMBLY_REPORTS		Directory	18.10.2012 ...	file (0444)	14 562
Acyrtosiphon_pisum		Directory	15.03.2012 ...	file (0444)	14 536
Aedes_aegypti		Directory	21.09.2010 ...	file (0444)	14 536
Ailuropoda_melanoleuca		Directory	06.09.2012 ...	file (0444)	14 536
Amphimedon_queenslandica		Directory	15.03.2012 ...	file (0444)	14 536
Anolis_carolinensis		Directory	15.03.2012 ...	file (0444)	14 536
Anopheles_gambiae		Directory	14.01.2008 ...	file (0444)	14 536
Apis_florea		Directory	13.04.2012 ...	file (0444)	14 536
Apis_mellifera		Directory	14.06.2011 ...	file (0444)	14 536
Arabidopsis_lyrata		Directory	11.05.2011 ...	file (0444)	14 536
Arabidopsis_thaliana		Directory	30.09.2008 ...	file (0444)	14 536
BACENDS		Directory	13.05.2010 ...	file (0444)	14 536
Bacteria		Directory	18.10.2012 ...	file (0444)	14 536
Bacteria_DRAFT		Directory	10.10.2012 ...	file (0444)	14 536
Bombus_impatiens		Directory	15.03.2012 ...	file (0444)	14 536
Bombus_terrestris		Directory	15.03.2012 ...	file (0444)	14 536

6 files and 88 directories. Total size: 15,4 MB

Server/Local file	Director	Remote file	Size	Priority	Status
-------------------	----------	-------------	------	----------	--------

Queued files Failed transfers Successful transfers

Queue: empty



# 1. RefSeq

# 1. ftp://ftp.ncbi.nih.gov/genomes/



Remote site: /genomes

- ? entrez
- ? epigenomics
- ? fa2htgs
- ? genbank
- ? gene
- ▶ ? genomes
- ? geo
- ▶ ? hapmap
- ? mmdb
- ? ncbi-asn1

Filename ^	Filesize	Filetype	Last modified	Permission	Owner/Gro
..					
ASSEMBLY_REPORTS		Directory	18.10.2012 ...	file (0444)	14 562
Acyrtosiphon_pisum		Directory	15.03.2012 ...	file (0444)	14 536
Aedes_aegypti		Directory	21.09.2010 ...	file (0444)	14 536
Ailuropoda_melanoleuca		Directory	06.09.2012 ...	file (0444)	14 536
Amphimedon_queenslandica		Directory	15.03.2012 ...	file (0444)	14 536
Anolis_carolinensis		Directory	15.03.2012 ...	file (0444)	14 536
Anopheles_gambiae		Directory	14.01.2008 ...	file (0444)	14 536
Apis_florea		Directory	13.04.2012 ...	file (0444)	14 536
Apis_mellifera		Directory	14.06.2011 ...	file (0444)	14 536
Arabidopsis_lyrata		Directory	11.05.2011 ...	file (0444)	14 536
Arabidopsis_thaliana		Directory	30.09.2008 ...	file (0444)	14 536
BACENDS		Directory	13.05.2010 ...	file (0444)	14 536
Bacteria		Directory	18.10.2012 ...	file (0444)	14 536
Bacteria_DRAFT		Directory	10.10.2012 ...	file (0444)	14 536
Bombus_impatiens		Directory	15.03.2012 ...	file (0444)	14 536
Bombus_terrestris		Directory	15.03.2012 ...	file (0444)	14 536

6 files and 88 directories. Total size: 15,4 MB

RefSeq



# 1. ftp://ftp.ncbi.nih.gov/genomes/Bacteria

Remote site: /genomes/Bacteria

- Amphimedon\_queenslandica
- Anolis\_carolinensis
- Anopheles\_gambiae
- Apis\_florea
- Apis\_mellifera
- Arabidopsis\_lyrata
- Arabidopsis\_thaliana
- ASSEMBLY\_REPORTS
- BACENDS
- Bacteria**

Filename ^	Filesize	Filetype	Last modified	Permission	Owner/Gro
Erythrobacter_litoralis_HTCC2594_uid58299		Directory	06.12.2010 ...	fle (0444)	14 536
<b>Escherichia_blattae_DSM_4481_uid165043</b>		<b>Directory</b>	<b>06.06.2012 ...</b>	<b>fle (0444)</b>	<b>14 536</b>
Escherichia_coli_042_uid161985		Directory	16.05.2012 ...	fle (0444)	14 536
Escherichia_coli_536_uid58531		Directory	06.12.2010 ...	fle (0444)	14 536
Escherichia_coli_55989_uid59383		Directory	06.12.2010 ...	fle (0444)	14 536
Escherichia_coli_ABU_83972_uid161975		Directory	05.05.2012 ...	fle (0444)	14 536
Escherichia_coli_APEC_O1_uid58623		Directory	06.12.2010 ...	fle (0444)	14 536
Escherichia_coli_ATCC_8739_uid58783		Directory	11.01.2011 ...	fle (0444)	14 536
Escherichia_coli_BL21_DE3_uid161947		Directory	17.05.2012 ...	fle (0444)	14 536
Escherichia_coli_BL21_DE3_uid161949		Directory	17.05.2012 ...	fle (0444)	14 536
Escherichia_coli_BW2952_uid59391		Directory	06.12.2010 ...	fle (0444)	14 536
Escherichia_coli_B_REL606_uid58803		Directory	06.12.2010 ...	fle (0444)	14 536
Escherichia_coli_CFT073_uid57915		Directory	06.12.2010 ...	fle (0444)	14 536
Escherichia_coli_DH1_uid161951		Directory	05.05.2012 ...	fle (0444)	14 536
Escherichia_coli_DH1_uid162051		Directory	16.05.2012 ...	fle (0444)	14 536
Escherichia_coli_E24377A_uid58395		Directory	06.12.2010 ...	fle (0444)	14 536
Escherichia_coli_ED1a_uid59379		Directory	06.12.2010 ...	fle (0444)	14 536

Selected 1 directory.

# 1. ftp://ftp.ncbi.nih.gov/genomes/Bacteria/ Escherichia\_coli\_042\_uid161985



Multiple  
replicons

Remote site: /genomes/Bacteria/Escherichia\_coli\_042\_uid161985

Escherichia\_blatteae\_DSM\_4481\_uid165043  
Escherichia\_coli\_042\_uid161985  
Escherichia\_coli\_536\_uid58531

Filename ^	Filesize	Filetype	Last modified	Permission	Owner/Gro
..					
<input type="checkbox"/> NC_017626.GeneMark-2.5m	1,2 MB	5m-file	16.05.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017626.GeneMarkHMM-2.6r	305,4 KB	6r-file	16.05.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017626.Glimmer3	206,5 KB	Glimmer...	16.05.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017626.Prodigal-2.50	1,1 MB	50-file	16.05.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017626.asn	23,8 MB	asn-file	14.09.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017626.faa	2,0 MB	faa-file	16.05.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017626.ffn	5,1 MB	ffn-file	16.05.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017626.fna	5,4 MB	fna-file	16.05.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017626.frn	45,2 KB	frn-file	16.05.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017626.gbk	20,6 MB	gbk-file	28.09.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017626.gff	5,4 MB	gff-file	16.05.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017626.ptt	398,5 KB	ptt-file	16.05.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017626.rnt	6,5 KB	rnt-file	16.05.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017626.rpt	281 B	rpt-file	16.05.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017626.val	10,1 MB	val-file	14.09.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017627.GeneMark-2.5m	28,0 KB	5m-file	16.05.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017627.GeneMarkHMM-2.6r	9,7 KB	6r-file	16.05.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017627.Glimmer3	7,2 KB	Glimmer...	16.05.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017627.Prodigal-2.50	30,3 KB	50-file	16.05.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017627.asn	553,7 KB	asn-file	28.09.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017627.faa	38,7 KB	faa-file	28.09.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017627.ffn	96,7 KB	ffn-file	16.05.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017627.fna	115,1 KB	fna-file	16.05.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017627.frn	166 B	frn-file	16.05.2012 ...	adfr (04...	14 536
<input type="checkbox"/> NC_017627.gbkbk	395,9 KB	gbkbk-file	28.09.2012 ...	adfr (04...	14 536

30 files. Total size: 76,9 MB

- Complete Genomes -

# 1. ftp://ftp.ncbi.nih.gov/genomes/Bacteria/ Escherichia\_coli\_042\_uid161985



Remote site: /genomes/Bacteria/Escherichia\_coli\_042\_uid161985

Escherichia\_blatteae\_DSM\_4481\_uid165043  
Escherichia\_coli\_042\_uid161985  
Escherichia\_coli\_536\_uid58531

Filename ^	Filesize	Filetype	Last modified	Permission	Owner/Gro
..					
NC_017626.GeneMark-2.5m	1,2 MB	5m-file	16.05.2012 ...	adfr (04...	14 536
NC_017626.GeneMarkHMM-2.6r	305,4 KB	6r-file	16.05.2012 ...	adfr (04...	14 536
NC_017626.Glimmer3	206,5 KB	Glimmer...	16.05.2012 ...	adfr (04...	14 536
NC_017626.Prodigal-2.50	1,1 MB	50-file	16.05.2012 ...	adfr (04...	14 536
NC_017626.asn	23,8 MB	asn-file	14.09.2012 ...	adfr (04...	14 536
NC_017626.faa	2,0 MB	faa-file	16.05.2012 ...	adfr (04...	14 536
NC_017626.ffn	5,1 MB	ffn-file	16.05.2012 ...	adfr (04...	14 536
NC_017626.fna	5,4 MB	fna-file	16.05.2012 ...	adfr (04...	14 536
NC_017626.frn	45,2 KB	frn-file	16.05.2012 ...	adfr (04...	14 536
NC_017626.gb	20,6 MB	gbk-file	28.09.2012 ...	adfr (04...	14 536
NC_017626.gff	5,4 MB	gff-file	16.05.2012 ...	adfr (04...	14 536
NC_017626.ptt	398,5 KB	ptt-file	16.05.2012 ...	adfr (04...	14 536
NC_017626.rnt	6,5 KB	rnt-file	16.05.2012 ...	adfr (04...	14 536
NC_017626.rpt	281 B	rpt-file	16.05.2012 ...	adfr (04...	14 536
NC_017626.val	10,1 MB	val-file	14.09.2012 ...	adfr (04...	14 536
NC_017627.GeneMark-2.5m	28,0 KB	5m-file	16.05.2012 ...	adfr (04...	14 536
NC_017627.GeneMarkHMM-2.6r	9,7 KB	6r-file	16.05.2012 ...	adfr (04...	14 536
NC_017627.Glimmer3	7,2 KB	Glimmer...	16.05.2012 ...	adfr (04...	14 536
NC_017627.Prodigal-2.50	30,3 KB	50-file	16.05.2012 ...	adfr (04...	14 536
NC_017627.asn	553,7 KB	asn-file	28.09.2012 ...	adfr (04...	14 536
NC_017627.faa	38,7 KB	faa-file	28.09.2012 ...	adfr (04...	14 536
NC_017627.ffn	96,7 KB	ffn-file	16.05.2012 ...	adfr (04...	14 536
NC_017627.fna	115,1 KB	fna-file	16.05.2012 ...	adfr (04...	14 536
NC_017627.frn	166 B	frn-file	16.05.2012 ...	adfr (04...	14 536
NC_017627.gb	395,9 KB	gbk-file	28.09.2012 ...	adfr (04...	14 536
30 files. Total size: 76,9 MB					

Multiple  
replicons

Multiple  
Project ids  
(uid)

Escherichia\_coli\_W\_uid162011  
Escherichia\_coli\_W\_uid162101

- Complete Genomes -

# File formats

- \*.asn: replicon in asn.1 format
- \*.faa: aa sequences of all CDS, multi-fasta
- \*.ffn: nt sequences of all CDS, multi-fasta
- \*.fna: replicon sequence, fasta
- \*.frn: nt sequences of all RNA genes, multi-fasta
- \*.gbk: replicon in genbank format
- \*GeneMark\*/Glimmer3/Prodigal: Genemark ab initio ORF finding results
- \*.gff: CDS/genes in GFF3 format (genome feature file)
- \*.ptt: CDS in ptt format (protein table file)
- \*.rnt: RNA genes table file
- \*.rpt: report file on features (e.g. acc#, GI, taxid, CDS count, gene count ...)
- \*.val: binary asn.1 format

# 1. ftp://ftp.ncbi.nih.gov/genomes/Bacteria



Remote site: /genomes/Bacteria

BACENDS

▼ Bacteria

- ? \_Cellvibrio\_gilvus\_ATCC\_13127\_uid68143
- ? \_Nostoc\_azollae\_\_0708\_uid49725
- ? Acaryochloris\_marina\_MBIC11017\_uid58167

Filename ^	Filesize	Filetype	Last modified	Permission	Owner/G
gamma_proteobacterium_HdN1_uid51635		Directory	06.12.2010 ...	file (0444)	14 536
halophilic_archaeon_DL31_uid72619		Directory	01.09.2011 ...	file (0444)	14 536
secondary_endosymbiont_of_Ctenarytaina_eucalypti_uid172737		Directory	20.08.2012 ...	file (0444)	14 536
uncultured_Termite_group_1_bacterium_phylotype_Rs_D17_uid5...		Directory	28.01.2011 ...	file (0444)	14 536
CLUSTERS	11 B	File	10.02.2012 ...	adfr (04...	14 536
ERR	313,5 KB	File	03.07.2012 ...	adfr (04...	14 536
NZ_AAEK01000176	18,0 KB	File	03.07.2012 ...	adfr (04...	14 536
ReadMe.txt	4,8 KB	txt-file	22.10.2010 ...	adfrw (0...	14 536
SameSpecies.gi	10,3 KB	gi-file	13.10.2012 ...	adfr (04...	14 536
all.GeneMark.tar.gz	421,2 MB	gz-file	18.10.2012 ...	adfr (04...	14 536
all.Glimmer3.tar.gz	94,9 MB	gz-file	18.10.2012 ...	adfr (04...	14 536
all.Prodigal.tar.gz	210,7 MB	gz-file	18.10.2012 ...	adfr (04...	14 536
all.asn.tar.gz	4,2 GB	gz-file	18.10.2012 ...	adfr (04...	14 536
all.faa.tar.gz	1,4 GB	gz-file	18.10.2012 ...	adfr (04...	14 536
all.ffn.tar.gz	2,1 GB	gz-file	18.10.2012 ...	adfr (04...	14 536
all.fna.tar.gz	2,3 GB	gz-file	18.10.2012 ...	adfr (04...	14 536
all.frn.tar.gz	8,5 MB	gz-file	18.10.2012 ...	adfr (04...	14 536
all.gbkb.tar.gz	6,3 GB	gz-file	18.10.2012 ...	adfr (04...	14 536
all.gff.tar.gz	507,0 MB	gz-file	18.10.2012 ...	adfr (04...	14 536
all.ptt.tar.gz	169,7 MB	gz-file	18.10.2012 ...	adfr (04...	14 536
all.rnt.tar.gz	2,5 MB	gz-file	18.10.2012 ...	adfr (04...	14 536
all.rpt.tar.gz	327,9 KB	gz-file	18.10.2012 ...	adfr (04...	14 536
all.val.tar.gz	3,8 GB	gz-file	18.10.2012 ...	adfr (04...	14 536

Selected 1 file. Total size: 6,3 GB

All Bacteria  
tarballs

- Complete Genomes -



# 1. [ftp://ftp.ncbi.nih.gov/genomes/Bacteria\\_DRAFT/Escherichia\\_coli\\_MS\\_115\\_1\\_uid50627](ftp://ftp.ncbi.nih.gov/genomes/Bacteria_DRAFT/Escherichia_coli_MS_115_1_uid50627)



Filename ^	Filesize	Filet
NZ_ADTL000000000.asn	17,0 KB	asn-fi
NZ_ADTL000000000.contig.asn.tgz	4,5 MB	tgz-fi
NZ_ADTL000000000.contig.faa.tgz	1,1 MB	tgz-fi
NZ_ADTL000000000.contig.ffn.tgz	1,5 MB	tgz-fi
NZ_ADTL000000000.contig.fna.tgz	1,6 MB	tgz-fi
NZ_ADTL000000000.contig.frn.tgz	8,9 KB	tgz-fi
NZ_ADTL000000000.contig.gbk.tgz	5,0 MB	tgz-fi
NZ_ADTL000000000.contig.gbs.tgz	32,7 KB	tgz-fi
NZ_ADTL000000000.contig.gff.tgz	412,3 KB	tgz-fi
NZ_ADTL000000000.contig.ptt.tgz	138,6 KB	tgz-fi
NZ_ADTL000000000.contig.rnt.tgz	4,2 KB	tgz-fi
NZ_ADTL000000000.contig.rpt.tgz	14,3 KB	tgz-fi
NZ_ADTL000000000.contig.val.tgz	3,7 MB	tgz-fi
NZ_ADTL000000000.gbk	5,3 KB	gbk-fi
NZ_ADTL000000000.rpt	264 B	rpt-fi
NZ_ADTL000000000.scaffold.asn.tgz	33,5 KB	tgz-fi
NZ_ADTL000000000.scaffold.faa.tgz	998,9 KB	tgz-fi
NZ_ADTL000000000.scaffold.ffn.tgz	1,5 MB	tgz-fi
NZ_ADTL000000000.scaffold.fna.tgz	1,6 MB	tgz-fi
NZ_ADTL000000000.scaffold.frn.tgz	8,5 KB	tgz-fi
NZ_ADTL000000000.scaffold.gbk.tgz	2,3 MB	tgz-fi
NZ_ADTL000000000.scaffold.gbs.tgz	19,5 KB	tgz-fi
NZ_ADTL000000000.scaffold.gff.tgz	5,3 KB	tgz-fi
NZ_ADTL000000000.scaffold.ptt.tgz	128,1 KB	tgz-fi
NZ_ADTL000000000.scaffold.rnt.tgz	3,7 KB	tgz-fi
NZ_ADTL000000000.scaffold.rpt.tgz	6,2 KB	tgz-fi

28 files. Total size: 24,1 MB

- Draft bacteria contain contigs or contigs + scaffolds
- Contigs/scaffolds packed in zipped tarballs
- Unzipped files are the master record
- Of course plasmids and chromosomes are mixed
- **Annotation either in scaffolds or contigs!!!**



## 2. Genbank

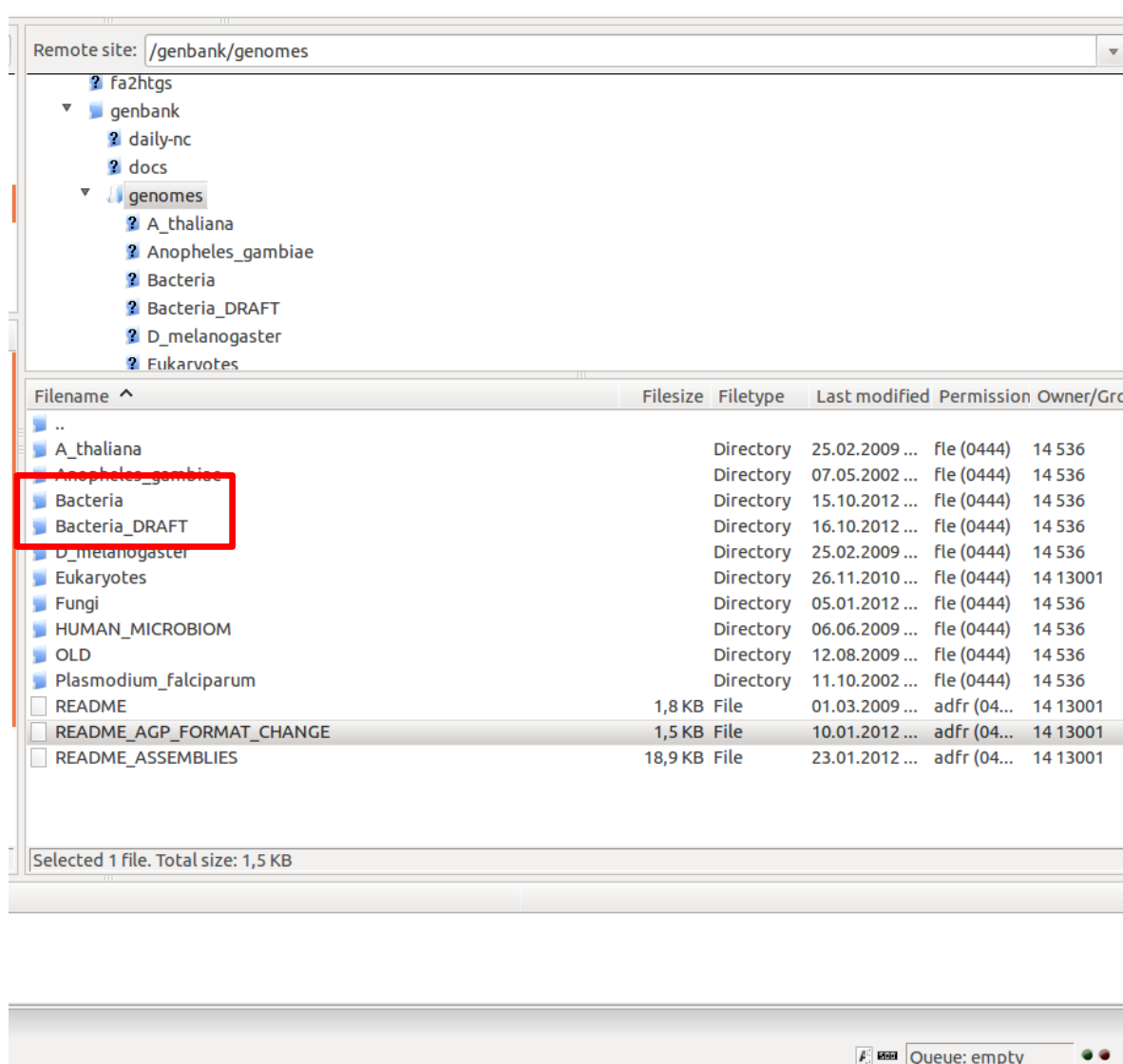
International Nucleotide Sequence  
Database Collaboration (INSDC)  
NCBI, EMBL, DDBJ

## 2. ftp://ftp.ncbi.nih.gov/genbank/ genomes/



Genbank

Same folder  
organisation as  
RefSeq, but no 'all'  
tarballs!



# RefSeq - Genbank numbers

Database	Bacteria	Bacteria_DRAFT	HUMAN_MICROBIOM/Bacteria
RefSeq	2105	1879	682
Genbank	2174	4026	899

# wget (- curl)

## Non-interactive network downloader

GNU Wget 1.11.4 Red Hat modified, a non-interactive network retriever.  
Usage: wget [OPTION]... [URL]...

Mandatory arguments to long options are mandatory for short options too.

### Startup:

-V, --version	display the version of Wget and exit.
-h, --help	print this help.
-b, --background	go to background after startup.
-e, --execute=COMMAND	execute a '.wgetrc'-style command.

### Logging and input file:

-o, --output-file=FILE	log messages to FILE.
-a, --append-output=FILE	append messages to FILE.
-d, --debug	print lots of debugging information.
-q, --quiet	quiet (no output).
-v, --verbose	be verbose (this is the default).
-nv, --no-verbose	turn off verboseness, without being quiet.
-i, --input-file=FILE	download URLs found in FILE.
-F, --force-html	treat input file as HTML.
-B, --base=URL	prepends URL to relative links in -F -i file.

### Download:

-t, --tries=NUMBER	set number of retries to NUMBER (0 unlimits).
--retry-connrefused	retry even if connection is refused.
-O, --output-document=FILE	write documents to FILE.
-nc, --no-clobber	skip downloads that would download to existing files.
-c, --continue	resume getting a partially-downloaded file.
--progress=TYPE	select progress gauge type.
-N, --timestamping	don't re-retrieve files unless newer than local.
-S, --server-response	print server response.
--spider	don't download anything.
-T, --timeout=SECONDS	set all timeout values to SECONDS.
--dns-timeout=SECS	set the DNS lookup timeout to SECS.
--connect-timeout=SECS	set the connect timeout to SECS.
--read-timeout=SECS	set the read timeout to SECS.
-w, --wait=SECONDS	wait SECONDS between retrievals.
--waitretry=SECONDS	wait 1..SECONDS between retries of a retrieval.
--random-wait	wait from 0...2*WAIT secs between retrievals.
--no-proxy	explicitly turn off proxy.
-Q, --quota=NUMBER	set retrieval quota to NUMBER.
--bind-address=ADDRESS	bind to ADDRESS (hostname or IP) on local host.
--limit-rate=RATE	limit download rate to RATE.
--no-dns-cache	disable caching DNS lookups.
--restrict-file-names=OS	restrict chars in file names to ones OS allows.
--ignore-case	ignore case when matching files/directories.
-4, --inet4-only	connect only to IPv4 addresses.
-6, --inet6-only	connect only to IPv6 addresses.
--prefer-family=FAMILY	connect first to addresses of specified family, one of IPv6, IPv4, or none.
--user=USER	set both ftp and http user to USER.
--password=PASS	set both ftp and http password to PASS.

### Directories:

-nd, --no-directories	don't create directories.
-x, --force-directories	force creation of directories.
-nH, --no-host-directories	don't create host directories.
--protocol-directories	use protocol name in directories.
-P, --directory-prefix=PREFIX	save files to PREFIX/...
--cut-dirs=NUMBER	ignore NUMBER remote directory components.

### HTTP options:

--http-user=USER	set http user to USER.
--http-password=PASS	set http password to PASS.
--no-cache	disallow server-cached data.
-E, --html-extension	save HTML documents with '.html' extension.

# wget

E.g. get all RefSeq E. coli genbanks and fastas:

```
wget -c -N -v -r -t 45 -nd -A  
      *.gbk, *.fna  
"ftp://ftp.ncbi.nih.gov/genomes/Bact  
  eria/Escherichia_coli*" -P .
```

# URLs

## RefSeq

- <ftp://ftp.ncbi.nih.gov/genomes/Bacteria/>
- [ftp://ftp.ncbi.nih.gov/genomes/Bacteria\\_DRAFT/](ftp://ftp.ncbi.nih.gov/genomes/Bacteria_DRAFT/)
- <ftp://ftp.ncbi.nih.gov/genomes/Fungi/>
- [ftp://ftp.ncbi.nih.gov/genomes/Fungi\\_DRAFT/](ftp://ftp.ncbi.nih.gov/genomes/Fungi_DRAFT/) - Currently empty
- <ftp://ftp.ncbi.nih.gov/genomes/Plasmids/>
- <ftp://ftp.ncbi.nih.gov/genomes/Viruses/>
- [ftp://ftp.ncbi.nih.gov/genomes/HUMAN\\_MICROBIOM/](ftp://ftp.ncbi.nih.gov/genomes/HUMAN_MICROBIOM/)
- [ftp://ftp.ncbi.nlm.nih.gov/genomes/GENOME\\_REPORTS/prokaryotes.txt](ftp://ftp.ncbi.nlm.nih.gov/genomes/GENOME_REPORTS/prokaryotes.txt)

## Genbank

- <ftp://ftp.ncbi.nih.gov/genbank/genomes/Bacteria/>
- [ftp://ftp.ncbi.nih.gov/genbank/genomes/Bacteria\\_DRAFT/](ftp://ftp.ncbi.nih.gov/genbank/genomes/Bacteria_DRAFT/)
- [ftp://ftp.ncbi.nih.gov/genbank/genomes/HUMAN\\_MICROBIOM/](ftp://ftp.ncbi.nih.gov/genbank/genomes/HUMAN_MICROBIOM/)
- <ftp://ftp.ncbi.nih.gov/genbank/genomes/Fungi>

[Include compressed 'all' tarballs](#)

#Organism/Name
BioProject Accession
BioProject ID
Group
SubGroup
Size (Mb)
GC%
Chromosomes/RefSeq
Chromosomes/INSDC
Plasmids/RefSeq
Plasmids/INSDC
WGS
Scaffolds
Genes
Proteins
Release Date
Modify Date
Status
Center

# Overview



ftp://ftp.ncbi.nlm.nih.gov/genomes/  
 GENOME\_REPORTS/prokaryotes.txt

#Organism/Name	BioProject Acc	BioProject ID	Group	SubGroup	Size (Mb)	GC%	Chromosomes/RefSeq	Chromosomes/INSDC	P
Escherichia coli O157:H7 str. FRIK966	PRJNA55561	55561	Proteobacteria	Gammaproteobacteria	537.691	50	-	-	-
Escherichia coli O157:H7 str. FRIK2000	PRJNA55943	55943	Proteobacteria	Gammaproteobacteria	540.869	50	-	-	-
Escherichia coli str. K-12 substr. MG1655	PRJNA57779	57779	Proteobacteria	Gammaproteobacteria	463.968	51	NC_000913.2	U00096.2	-
Escherichia coli O157:H7 str. Sakai	PRJNA57781	57781	Proteobacteria	Gammaproteobacteria	559.448	####	NC_002695.1	BA000007.2	N
Escherichia coli O157:H7 str. EDL933	PRJNA57831	57831	Proteobacteria	Gammaproteobacteria	562.052	####	NC_002655.2	AE005174.2	N
Escherichia coli CFT073	PRJNA57915	57915	Proteobacteria	Gammaproteobacteria	523.143	51	NC_004431.1	AE014075.1	-
Escherichia coli HS	PRJNA58393	58393	Proteobacteria	Gammaproteobacteria	464.354	51	NC_009800.1	CP000802.1	-


# EMBL – EBI ftp server

<http://www.ebi.ac.uk/FTP/>

EMBL-EBI   [Terms of Use](#) [Privacy](#) [Cookies](#)


[Databases](#) [Tools](#) [Research](#) [Training](#) [Industry](#) [About Us](#) [Help](#) [Site Index](#)  

- EBI FTP Server
- Download Help files
- Database Repository
- Software Repository
- Submissions

**UniProt** 


The premier centralised resource for protein sequences and functional information.

- [README](#)
- [Release Notes](#)
- [Download](#)

**UniProtKB/ Swiss-Prot** 


An annotated protein sequence database.

- [README](#)
- [Release Notes](#)
- [Download](#)

**UniProtKB/ TrEMBL** 


A computer generated protein database enriched with automated classification and annotation.

- [README](#)
- [Release Notes](#)
- [Download](#)

**EMBL** 

Europe's primary nucleotide sequence resource.

- [Release Notes](#)


**InterPro** 














A database of integrated documentation resources for protein families, domains and functional sites.

- [README](#)
- [Release Notes](#)
- [Download](#)

EBI > Downloads

### Downloads from the EBI FTP Server

The European Bioinformatics Institute (EBI) FTP server provides a comprehensive archive of databases and software in the field of bioinformatics. Below are the main directory links of this server.  downloads

Directory	Explanation
<b>Main Directories</b>	
<a href="#">databases</a>	The Database Repository
<a href="#">software</a>	The Software Repository
<a href="#">help</a>	Help Files
<b>Data Resources</b>	
<a href="#">1000 genomes</a>	1000 genomes project sequencing data FTP site
<a href="#">ArrayExpress</a> 	ArrayExpress Database FTP Site
<a href="#">ATD</a> 	ATD Database FTP Site
<a href="#">BioModels Database</a> 	Database of Annotated Published Models FTP Site
<a href="#">chEBI</a>	Dictionary and ontology of chemical entities FTP site
<a href="#">chEMBL</a>	Database of bioactive drug-like small molecules FTP site
<a href="#">dssp</a> 	DSSP Database FTP Site
<a href="#">dgvva</a>	Repository of genomic variants FTP Site
<a href="#">embl</a> 	EMBL Nucleotide Database FTP Site
<a href="#">fastafiles</a>	ENA, UniProt and Patent sequence data FTP site
<a href="#">fssp</a> 	FSSP Database FTP Site
<a href="#">emdb</a>	Electron microscopy density maps of macromolecular complexes and subcellular structures.
<a href="#">Ensembl</a>	Ensembl FTP site
<a href="#">Ensembl Genomes</a>	Ensembl Genomes FTP site
<a href="#">goa</a> 	GOA Database FTP Site
<a href="#">hssp</a> 	HSPP Database FTP Site
<a href="#">Integr8</a>	Genome_reviews data FTP site
<a href="#">interpro</a> 	InterPro Database FTP Site
<a href="#">Intenz</a>	Enzyme data FTP site
<a href="#">IntAct</a> 	IntAct data in PSI MI XML format
<a href="#">Irrex</a>	Stable genomic reference sequences.
<a href="#">Microcosm</a>	Predicted targets for microRNA sequences.
<a href="#">HMRShtDB</a> 	NMRShiftDB is a NMR database for organic structures and their nuclear magnetic resonance spectra.
<a href="#">Patentdata</a>	EPO data FTP site
<a href="#">PDBe</a> 	PDBe Database FTP Site
<a href="#">PRIDE</a>	PRIDE proteomics data FTP site
<a href="#">pdb_finder</a>	Mirror of PDBFINDER ( <a href="http://swift.cmbi.ru.nl/gv/pdbfinder/">http://swift.cmbi.ru.nl/gv/pdbfinder/</a> ) from CMBI ( <a href="ftp://ftp.cmbi.kun.nl/pub/molbio/data/pdbfinder/">ftp://ftp.cmbi.kun.nl/pub/molbio/data/pdbfinder/</a> ).
<a href="#">RESID</a> 	RESID Database FTP Site
<a href="#">RHEA</a>	Biochemical reactions data FTP site

- Use for faster downloads, e.g. whole 'nr' database

<ftp://ftp.ebi.ac.uk/pub/databases/>