NCBI
ftp server

for microbial genomes

Andreas Leimbach

18-10-2012

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 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

Software engineering
tools, R&D and databases

Education
teaching resources and on-line tutorials

Contact information
how to reach us

Major resources available by ftp (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).

▶ **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.

▶ **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/

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):

▶ 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).

▶ 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
Canis_familiaris		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/

FileZilla *The free FTP solution*

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

PayPal
DONATE

Sponsors:
Supported by
artmotion

NDC HOST

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/

1. RefSeq

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

RefSeq

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 ...	fle (0444)	14 562
Acyrthosiphon_pisum		Directory	15.03.2012 ...	fle (0444)	14 536
Aedes_aegypti		Directory	21.09.2010 ...	fle (0444)	14 536
Ailuropoda_melanoleuca		Directory	06.09.2012 ...	fle (0444)	14 536
Amphimedon_queenslandica		Directory	15.03.2012 ...	fle (0444)	14 536
Anolis_carolinensis		Directory	15.03.2012 ...	fle (0444)	14 536
Anopheles_gambiae		Directory	14.01.2008 ...	fle (0444)	14 536
Apis_florea		Directory	13.04.2012 ...	fle (0444)	14 536
Apis_mellifera		Directory	14.06.2011 ...	fle (0444)	14 536
Arabidopsis_lyrata		Directory	11.05.2011 ...	fle (0444)	14 536
Arabidopsis_thaliana		Directory	30.09.2008 ...	fle (0444)	14 536
BACENDS		Directory	13.05.2010 ...	fle (0444)	14 536
Bacteria		Directory	18.10.2012 ...	fle (0444)	14 536
Bacteria_DRAFT		Directory	10.10.2012 ...	fle (0444)	14 536
Bombus_impatiens		Directory	15.03.2012 ...	fle (0444)	14 536
Bombus_terrestris		Directory	15.03.2012 ...	fle (0444)	14 536

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

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

- Complete Genomes -

1. ftp://ftp.ncbi.nih.gov/genomes/**Bacteria**/ Escherichia_coli_042_uid161985

Multiple replicons

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.gbk		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.gbk		395,9 KB	gbk-file	28.09.2012 ...	adfr (04...	14 536

Remote site: /genomes/Bacteria/Escherichia_coli_042_uid161985

? Escherichia_blattae_DSM_4481_uid165043
Escherichia_coli_042_uid161985
? Escherichia_coli_536_uid58531

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_blattae_DSM_4481_uid165043
	Escherichia_coli_042_uid161985
?	Escherichia_coli_536_uid58531

Multiple replicons

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.gbk	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.gbk	395,9 KB	gbk-file	28.09.2012 …	adfr (04…	14 536

30 files. Total size: 76,9 MB

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 ...	fle (0444)	14 536
halophilic_archaeon_DL31_uid72619		Directory	01.09.2011 ...	fle (0444)	14 536
secondary_endosymbiont_of_Ctenarytaina_eucalypti_uid172737		Directory	20.08.2012 ...	fle (0444)	14 536
uncultured_Termite_group_1_bacterium_phylotype_Rs_D17_uid5...		Directory	28.01.2011 ...	fle (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.gbk.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

Remote site: /genomes/Bacteria_DRAFT/Escherichia_coli_MS_115_1_uid50627

- Escherichia_coli_MS_107_1_uid50373
- Escherichia_coli_MS_115_1_uid50627
- Escherichia_coli_MS_116_1_uid50635

Filename ^	Filesize	Filet
NZ_ADTL00000000.asn	17,0 KB	asn-fi
NZ_ADTL00000000.contig.asn.tgz	4,5 MB	tgz-fi
NZ_ADTL00000000.contig.faa.tgz	1,1 MB	tgz-fi
NZ_ADTL00000000.contig.ffn.tgz	1,5 MB	tgz-fi
NZ_ADTL00000000.contig.fna.tgz	1,6 MB	tgz-fi
NZ_ADTL00000000.contig.frn.tgz	8,9 KB	tgz-fi
NZ_ADTL00000000.contig.gbk.tgz	5,0 MB	tgz-fi
NZ_ADTL00000000.contig.gbs.tgz	32,7 KB	tgz-fi
NZ_ADTL00000000.contig.gff.tgz	412,3 KB	tgz-fi
NZ_ADTL00000000.contig.ptt.tgz	138,6 KB	tgz-fi
NZ_ADTL00000000.contig.rnt.tgz	4,2 KB	tgz-fi
NZ_ADTL00000000.contig.rpt.tgz	14,3 KB	tgz-fi
NZ_ADTL00000000.contig.val.tgz	3,7 MB	tgz-fi
NZ_ADTL00000000.gbk	5,3 KB	gbk-f
NZ_ADTL00000000.rpt	264 B	rpt-fi
NZ_ADTL00000000.scaffold.asn.tgz	33,5 KB	tgz-fi
NZ_ADTL00000000.scaffold.faa.tgz	998,9 KB	tgz-fi
NZ_ADTL00000000.scaffold.ffn.tgz	1,5 MB	tgz-fi
NZ_ADTL00000000.scaffold.fna.tgz	1,6 MB	tgz-fi
NZ_ADTL00000000.scaffold.frn.tgz	8,5 KB	tgz-fi
NZ_ADTL00000000.scaffold.gbk.tgz	2,3 MB	tgz-fi
NZ_ADTL00000000.scaffold.gbs.tgz	19,5 KB	tgz-fi
NZ_ADTL00000000.scaffold.gff.tgz	5,3 KB	tgz-fi
NZ_ADTL00000000.scaffold.ptt.tgz	128,1 KB	tgz-fi
NZ_ADTL00000000.scaffold.rnt.tgz	3,7 KB	tgz-fi
NZ_ADTL00000000.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

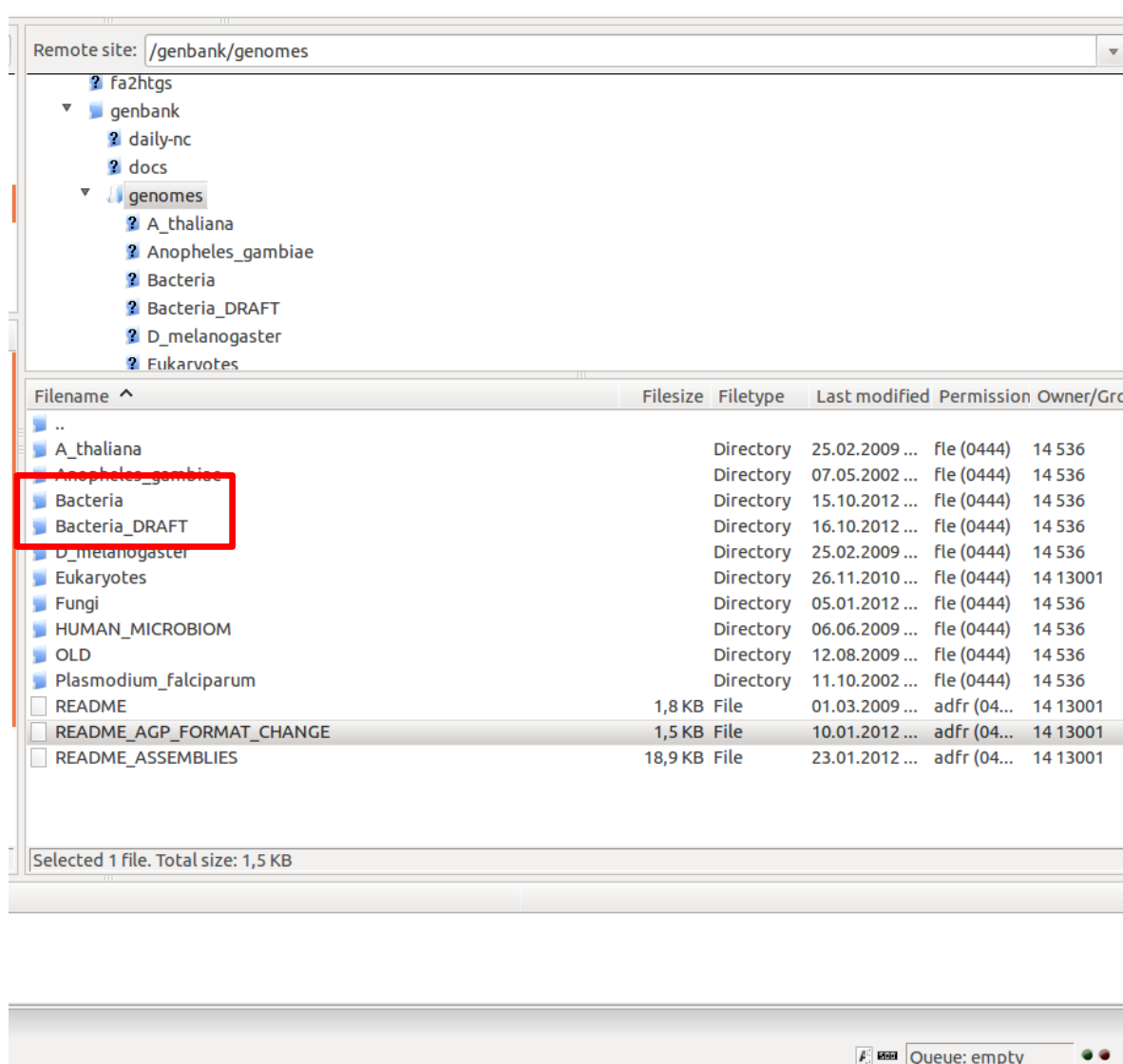- **Annotation either** in scaffolds or contigs!!!

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/Fungi/
- 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.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/HUMAN_MICROBIOM/
- ftp://ftp.ncbi.nih.gov/genbank/genomes/Fungi

Include compressed 'all' tarballs

Overview

#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

ftp://ftp.ncbi.nlm.nih.gov/genomes/
GENOME_REPORTS/prokaryotes.txt

#Organism/Name	BioProject Acce	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

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

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