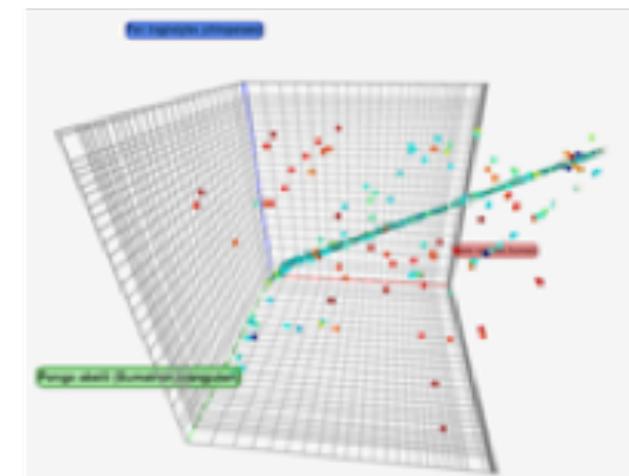
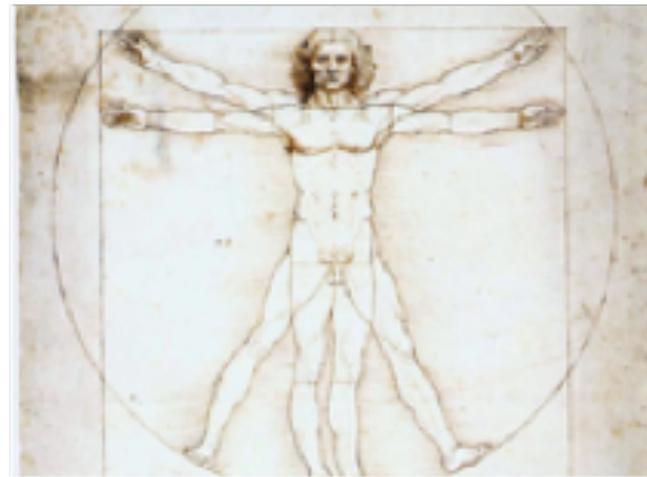
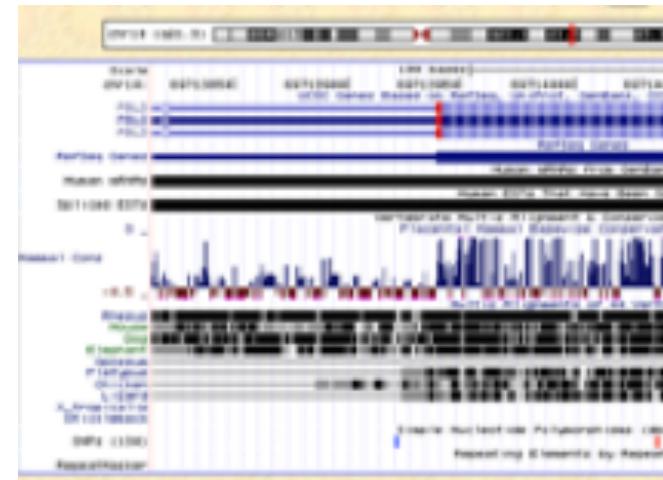


Computational Genomics

BLAST In Galaxy



BLAST in Galaxy

Output formats of BLAST

- **First Option:**

- Because Galaxy focuses on processing tabular data, the default output of this tool is tabular
- The 12 column BLAST+ tabular output contains:

Column	NCBI Name	Description
1	qseqid	Query Seq-id (ID of your sequence)
2	sseqid	Subject Seq-id (ID of the database hit)
3	pident	Percentage of identical matches
4	length	Alignment length
5	mismatch	Number of mismatches
6	gapopen	Number of gap openings
7	qstart	Start of alignment in query
8	qend	End of alignment in query
9	sstart	Start of alignment in subject (database hit)
10	send	End of alignment in subject (database hit)
11	evalue	Expectation value (E-value)
12	bitscore	Bit score

BLAST in Galaxy

Output formats of BLAST

- **Second Option:**

- The BLAST+ tools can optionally output additional columns of information, but this takes longer to calculate
- Many commonly used extra columns are included by selecting the extended tabular output
- The extra columns are included after the standard 12 columns
- This is so that you can write workflow filtering steps that accept either the 12 or 25 column tabular BLAST output
- Galaxy now uses this extended 25 column output by default:

Column	NCBI Name	Description
13	sallseqid	All subject Seq-id(s), separated by a ‘;’
14	score	Raw score
15	nident	Number of identical matches
16	positive	Number of positive-scoring matches
17	gaps	Total number of gaps
18	ppos	Percentage of positive-scoring matches
19	qframe	Query frame
20	sframe	Subject frame
21	qseq	Aligned part of query sequence
22	sseq	Aligned part of subject sequence
23	qlen	Query sequence length
24	slen	Subject sequence length
25	salltitles	All subject title(s), separated by a ‘<>’

BLAST in Galaxy

Output formats of BLAST

- **Third Option:**
 - To customize the tabular output by selecting which columns you want, from the standard set of 12, the default set of 25, or any of the additional columns BLAST+ offers (including species name)
- **Fourth Option:**
 - Request BLAST XML output, which is designed to be parsed by another program, and is understood by some Galaxy tools
- **Other Options:**
 - You can also choose several plain text or HTML output formats which are designed to be read by a person (not by another program)
 - The HTML versions use basic webpage formatting and can include links to the hits on the NCBI website
 - The pairwise output (the default on the NCBI BLAST website) shows each match as a pairwise alignment with the query
 - The two query anchored outputs show a multiple sequence alignment between the query and all the matches, and differ in how insertions are shown (marked as insertions or with gap characters added to the other sequences)

BLAST in Galaxy

Download FileZilla

The screenshot shows the official FileZilla website at <https://filezilla-project.org>. The page features a large red banner with the FileZilla logo and the tagline "The free FTP solution". On the left, a sidebar contains links for Home, FileZilla (Features, Screenshots, Download, Documentation, FileZilla Pro), FileZilla Server (Download), Community (Forum, Wiki), General (FAQ, Support, Contact, License, Privacy Policy, Trademark Policy), Development (Source code, Nightly builds, Translations, Version history, Changelog, Issue tracker), and Other projects (libfilezilla, Octochess). The main content area includes an advertisement for Inmotion hosting, a section titled "Overview" with information about the free client, server, forums, and documentation, and a "Quick download links" section with buttons for "Download FileZilla Client" and "Download FileZilla Server". Below these are news items for FileZilla Server 1.3.0, FileZilla Client 3.58.0, and FileZilla Client 3.58.0-rc1, each with a list of bugfixes and minor changes. A sidebar on the right promotes the "CODE TO CLOUD VIRTUAL SUMMIT" and the "PRISMA CLOUD bridgecrew".

Download FileZilla

FileZilla - The free FTP solution

FileZilla The free FTP solution

Overview

Welcome to the homepage of FileZilla®, the free FTP solution. The *FileZilla Client* not only supports FTP, but also FTP over TLS (FTPS) and SFTP. It is open source software distributed free of charge under the terms of the GNU General Public License.

We are also offering *FileZilla Pro*, with additional protocol support for WebDAV, Amazon S3, Backblaze B2, Dropbox, Microsoft OneDrive, Google Drive, Microsoft Azure Blob and File Storage, and Google Cloud Storage.

Last but not least, *FileZilla Server* is a free open source FTP and FTPS Server.

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](#)  [Download FileZilla Server](#) 

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

News

2022-02-14 - FileZilla Server 1.3.0 released

New features:

- Configuration wizard to setup passive mode
- Linux: Warn if sysctl knob kernel.yama.ptrace_scope is 0
- Linux: Refuse to run if sysctl knob fs.protected_hardlinks is 0

Bugfixes and minor changes:

- Admin UI: Rejecting a certificate fingerprint prompt no longer triggers automatic reconnections

2022-02-11 - FileZilla Client 3.58.0 released

Bugfixes and minor changes:

- More detailed error messages if an XML file known to exist cannot be opened

2022-02-03 - FileZilla Client 3.58.0-rc1 released

New features:

- Negotiate custom ALPN with FileZilla Server >=1.3.0 during TLS handshake to save some roundtrips during connection establishment

Advertisement:  Buy a Power Plan and Save Over \$200 Offer Available on Shared Hosting 11/22 - 12/7

Advertisement:  PRISMA CLOUD bridgecrew REGISTER NOW

Don't miss the global DevSecOps summit made by practitioners for practitioners. Coming up on March 23 & 24.

BLAST in Galaxy

Open FileZilla

FileZilla

Host: Username: Password: Port: Quickconnect

Local site: /etc/

- bin
- cores
- > dev
- > etc
- home
- > opt
- > private
- sbin
- > tmp
- > usr
- > var

Filename	Filesize	Filetype	Last modified	Permissions	Owner
..					
apache2		Directory	02/06/2022 15:2...		
asl		Directory	02/06/2022 15:2...		
cups		Directory	02/11/2022 11:0...		
defaults		Directory	02/06/2022 15:2...		
emond.d		Directory	02/06/2022 15:2...		
manpaths.d		Directory	02/11/2022 11:0...		
newsyslog.d		Directory	02/06/2022 15:2...		
openldap		Directory	02/06/2022 15:2...		
pam.d		Directory	02/06/2022 15:2...		
paths.d		Directory	02/11/2022 11:0...		
periodic		Directory	02/06/2022 15:2...		
pt.anchors		Directory	02/06/2022 15:2...		
postfix		Directory	02/06/2022 15:2...		
ppp		Directory	02/06/2022 15:2...		
racoon		Directory	02/06/2022 15:2...		
security		Directory	02/06/2022 15:2...		
snmp		Directory	02/06/2022 15:2...		
ssh		Directory	02/11/2022 11:0...		
ssl		Directory	02/06/2022 15:2...		
sudoers.d		Directory	02/06/2022 15:2...		
uucp		Directory	02/06/2022 15:2...		
wfs		Directory	02/06/2022 15:2...		

59 files and 22 directories. Total size: 821,780 bytes

Remote site:

Not connected to any server

Not connected.

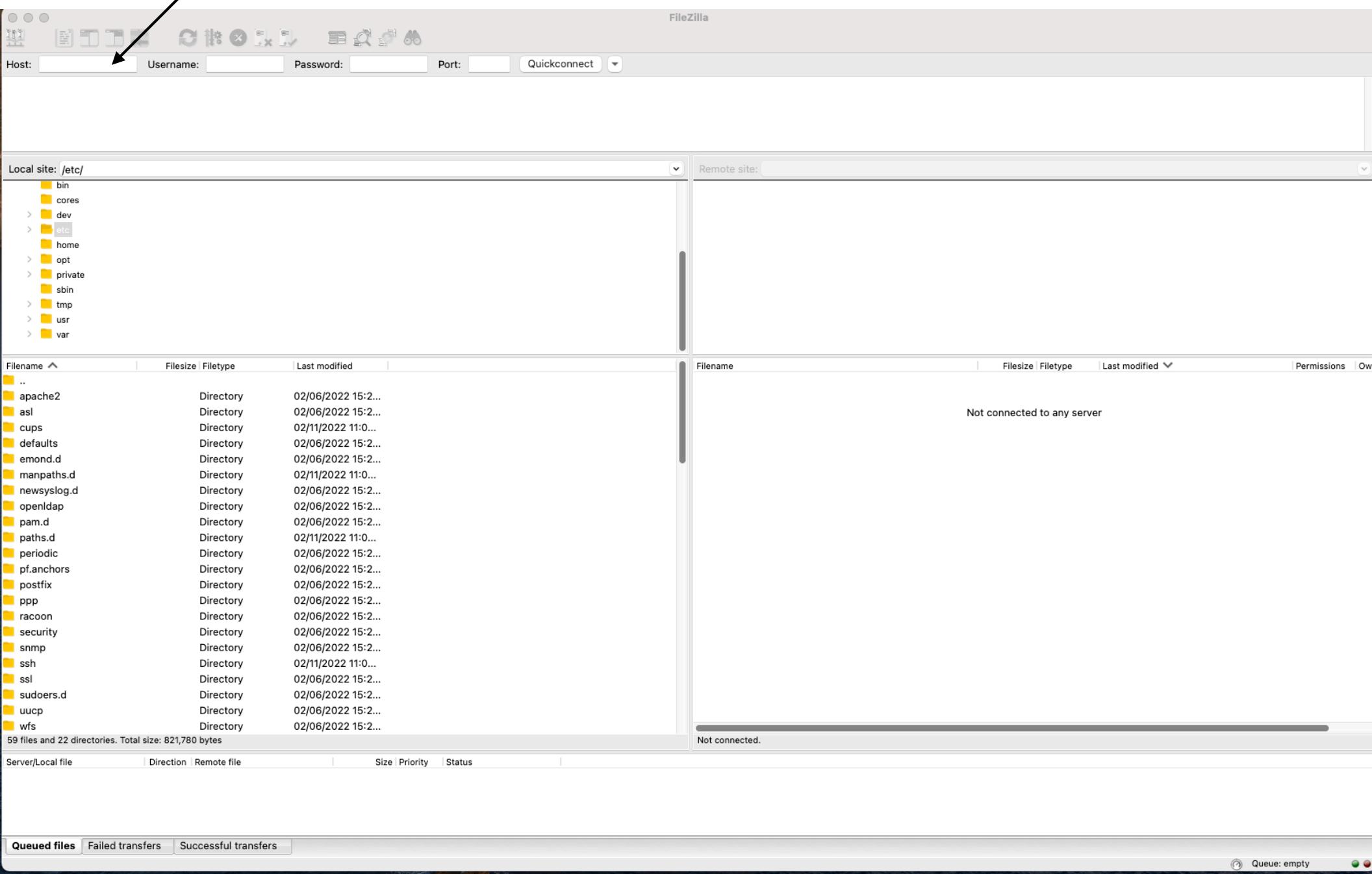
Server/Local file | Direction | Remote file | Size | Priority | Status

Queued files Failed transfers Successful transfers

Queue: empty

BLAST in Galaxy

Enter: ftp://ftp.ensembl.org/pub/release-105/fasta/homo_sapiens/pep/



BLAST in Galaxy

Go to your Downloads directory

The screenshot shows the FileZilla interface with two panes. The left pane represents the Local site (User's computer) and the right pane represents the Remote site (ftp://ftp.ensembl.org). A large black arrow points from the text "Go to your Downloads directory" down to the Local site pane, specifically highlighting the "Downloads" folder.

Local site: /Users/raramayo/Downloads/

- .cups
- .emacs.d
- .gnupg
- .igv
- .local
- .ssh
- .zsh_sessions
- Applications
- Desktop
- Documents
- Downloads**
- GIT

Filename	Filesize	Filetype	Last modified	Permissions	Owner
..					
.DS_Store	6,148	File	02/2022 13:4...	-rw-r--r--	ftp
.Rhistory	5,170	File	03/12/2021 09:5...	-rw-r--r--	ftp
.localized	0	File	12/11/2021 18:15...	-rw-r--r--	ftp

3 files. Total size: 11,318 bytes

Remote site: /pub/release-105/fasta/homo_sapiens/pep

- /
- pub
- release-105
- fasta
- homo_sapiens
- pep

Filename	Filesize	Filetype	Last modified	Permissions	Owner
..					
CHECKSUMS	116	File	06/20/2021	-rw-r--r--	ftp
README	2,432	File	06/18/2021	-rw-r--r--	ftp
Homo_sapiens.GRCh38.pep.all.fa.gz	14,376,754	gz-file	06/18/2021	-rw-r--r--	ftp
Homo_sapiens.GRCh38.pep.ab initio.fa.gz	13,738,129	gz-file	06/18/2021	-rw-r--r--	ftp

4 files. Total size: 28,117,431 bytes

Server/Local file | Direction | Remote file | Size | Priority | Status |

Queued files Failed transfers Successful transfers

Queue: empty

BLAST in Galaxy

Select the desired proteins file

The screenshot shows the FileZilla interface with two panes. The left pane (Local site) shows the contents of the user's Downloads folder, which includes .emacs.d, .gnupg, .igv, .local, .ssh, .zsh_sessions, Applications, Desktop, Documents, Downloads (selected), and GIT. The right pane (Remote site) shows the directory structure on ftp.ensembl.org at /pub/release-105/fasta/homo_sapiens/pep. A large black arrow points from the text "Select the desired proteins file" to the pep file in the remote directory. The local file list shows 3 files total size 11,318 bytes. The remote file list shows 4 files total size 28,117,431 bytes.

Host: ftp.ensembl.org Username: Password: Port: Quickconnect

Status: Resolving address of ftp.ensembl.org
Status: Connecting to 193.62.193.139:21...
Status: Connection established, waiting for welcome message...
Status: Insecure server, it does not support FTP over TLS.
Status: Logged in
Status: Retrieving directory listing of "/pub/release-105/fasta/homo_sapiens/pep"...
Status: Directory listing of "/pub/release-105/fasta/homo_sapiens/pep" successful

Local site: /Users/raramayo/Downloads/

Remote site: /pub/release-105/fasta/homo_sapiens/pep

Filename	Filesize	Filetype	Last modified	Permissions	Owne
..					
CHECKSUMS	116	File	06/20/2021	-rw-r--r--	ftp
README	2,432	File	06/18/2021	-rw-r--r--	ftp
Homo_sapiens.GRCh38.pep.all.fa.gz	14,376,754	gz-file	06/18/2021	-rw-r--r--	ftp
Homo_sapiens.GRCh38.pep.ab initio.fa.gz	13,738,129	gz-file	06/18/2021	-rw-r--r--	ftp

3 files. Total size: 11,318 bytes

4 files. Total size: 28,117,431 bytes

Server/Local file | Direction | Remote file | Size | Priority | Status |

Queued files Failed transfers Successful transfers Queue: empty

BLAST in Galaxy

Download file

ftp.ensembl.org - FileZilla

Error: Disconnected from server

Status: Resolving address of ftp.ensembl.org

Status: Connecting to 193.62.193.139:21...

Status: Connection established, waiting for welcome message...

Status: Insecure server, it does not support FTP over TLS.

Status: Logged in

Status: Starting download of /pub/release-105/fasta/homo_sapiens/pep/Homo_sapiens.GRCh38.pep.all.fa.gz

Local site: /Users/raramayo/Downloads/

- .cups
- .emacs.d
- .gnupg
- .igv
- .local
- .ssh
- .zsh_sessions
- Applications
- Desktop
- Documents
- Downloads
- GIT

Filename	Filesize	Filetype	Last modified	Permissions	Ow
..					
.DS_Store	6,148	File	02/21/2022 13:4...		
.Rhistory	5,170	File	03/12/2021 09:5...		
.localized	0	File	12/11/2021 18:15...		

3 files. Total size: 11,318 bytes

Remote site: /pub/release-105/fasta/homo_sapiens/pep

- /
- pub
- release-105
- fasta
- homo_sapiens
- pep

Filename	Filesize	Filetype	Last modified	Permissions	Ow
..					
CHECKSUMS	116	File	06/20/2021	-rw-r--r--	ftp
README	2,432	File	06/18/2021	-rw-r--r--	ftp
Homo_sapiens.GRCh38.pep.all.fa.gz	14,376,754	gz-file	06/18/2021	-rw-r--r--	ftp
Homo_sapiens.GRCh38.pep.ab initio.fa.gz	13,738,129	gz-file	06/18/2021	-rw-r--r--	ftp

Selected 1 file. Total size: 14,376,754 bytes

Server/Local file | Direction | Remote file | Size | Priority | Status

ftp.ensembl.org | /Users/raramayo/Down. | <<-- /pub/release-105/fasta/homo_sapiens/pep/Homo_sapiens.GRCh38.pep.all.fa.gz | 14,376,754 | Normal | Transferring
00:00:06 elapsed 00:00:03 left 76.5% 11,010,048 bytes (2.2 MiB/s)

Queued files (1) Failed transfers Successful transfers

Queue: 13.8 MiB

BLAST in Galaxy

Download file

ftp.ensembl.org - FileZilla

Host: ftp.ensembl.org Username: Password: Port: Quickconnect

Status: Resolving address of ftp.ensembl.org
Status: Connecting to 193.62.193.139:21...
Status: Connection established, waiting for welcome message...
Status: Insecure server, it does not support FTP over TLS.
Status: Logged in
Status: Starting download of /pub/release-105/fasta/homo_sapiens/pep/Homo_sapiens.GRCh38.pep.all.fa.gz
Status: File transfer successful, transferred 14,376,754 bytes in 7 seconds

Local site: /Users/raramayo/Downloads/

- .emacs.d
- .gnupg
- .igv
- .local
- .ssh
- .zsh_sessions
- Applications
- Desktop
- Documents
- Downloads
- GIT

Filename	Filesize	Filetype	Last modified	Permissions	Owner
..					
.DS_Store	6,148	File	02/21/2022 13:4...		
.Rhistory	5,170	File	03/12/2021 09:5...		
.localized	0	File	12/11/2021 18:15...		
Homo_sapiens.GRCh...	14,376,754	gz-file	02/21/2022 17:1...		

4 files. Total size: 14,388,072 bytes

Selected 1 file. Total size: 14,376,754 bytes

Remote site: /pub/release-105/fasta/homo_sapiens/pep

- /
- pub
- release-105
- fasta
- homo_sapiens
- pep

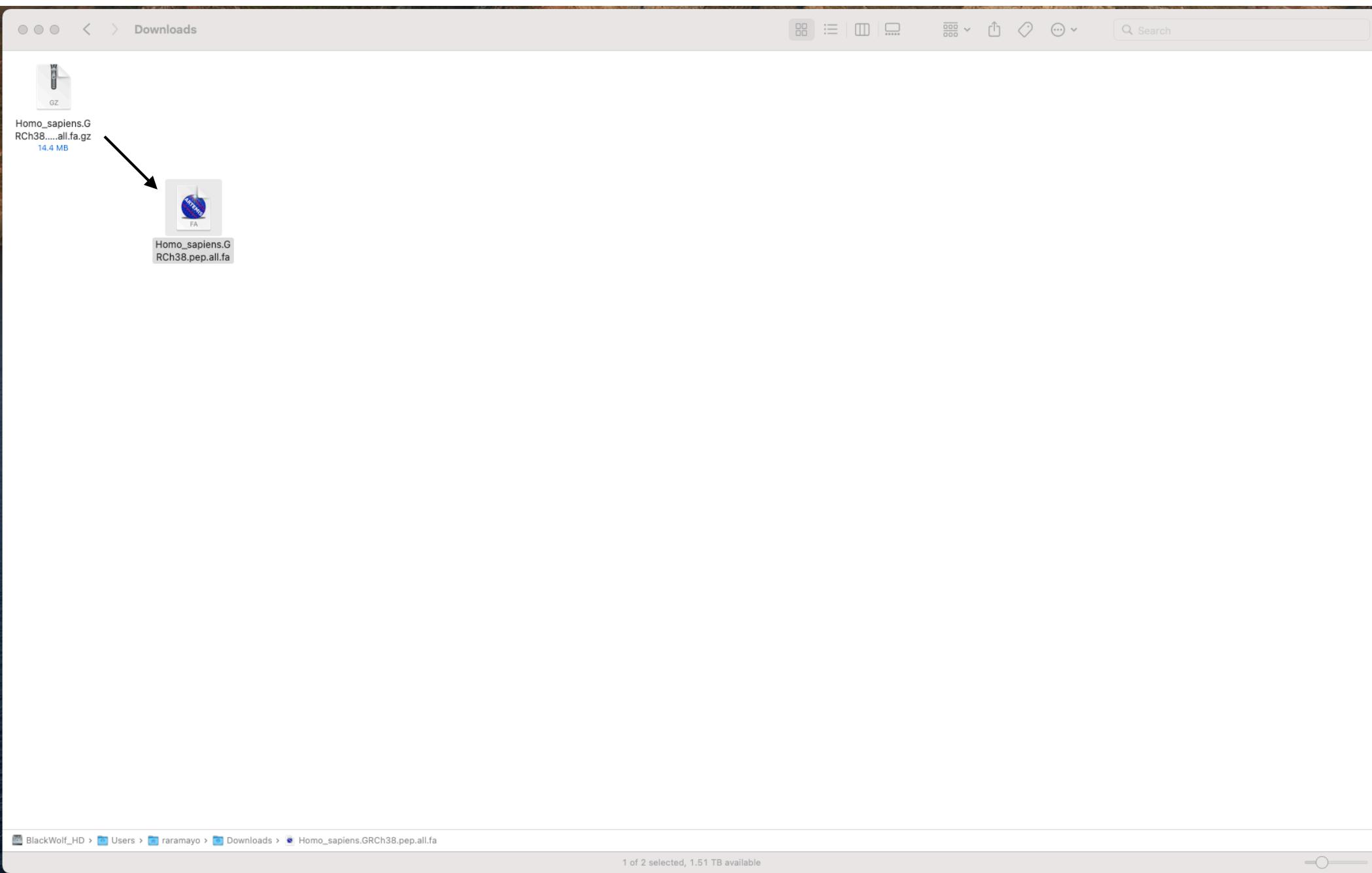
Filename	Filesize	Filetype	Last modified	Permissions	Owner
..					
CHECKSUMS	116	File	06/20/2021	-rw-r--r--	ftp
README	2,432	File	06/18/2021	-rw-r--r--	ftp
Homo_sapiens.GRCh38.pep.all.fa.gz	14,376,754	gz-file	06/18/2021	-rw-r--r--	ftp
Homo_sapiens.GRCh38.pep.ab initio.fa.gz	13,738,129	gz-file	06/18/2021	-rw-r--r--	ftp

Server/Local file | Direction | Remote file | Size | Priority | Status

Queued files Failed transfers Successful transfers (1) Queue: empty

BLAST in Galaxy

Decompress file



BLAST in Galaxy

Upload file into Kaiser

Galaxy https://galaxy-terra.hprc.tamu.edu/kaiser

HPRC Kaiser Galaxy

Tools

- search tools

HPRC

- Get Data
- Send Data
- Collection Operations
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- Fetch Alignments/Sequences
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- FASTQ Quality Control
- Mapping
- NCBI BLAST+
- Assembly
- Annotation
- FASTA/FASTQ
- Datamash
- EMBOSS
- Multiple Alignments
- Picard
- BED

Download from web or upload from disk

Regular Composite Collection Rule-based

Drop files here

Type (set all): Auto-detect Q Genome (set all): Additional ...

Choose local files Choose FTP files Paste/Fetch data Start Select Pause Reset Close

COVID-19 related research on Galaxy: training, tutorials, documents

Current known issues: Some tools cannot be removed from favorites. Choose your favorites wisely.

Uploading data from your computer from Galaxy Project

History

Lecture16 (empty)

This history is empty. You can load your own data or get data from an external source

BLAST in Galaxy

Upload file into Kaiser

The screenshot shows the HPRC Kaiser Galaxy web interface. A modal dialog box titled "Download from web or upload from disk" is open in the center. The dialog has tabs for "Regular", "Composite", "Collection", and "Rule-based". On the left, there's a sidebar with "Favorites" containing "Applications", "Downloads" (which is selected), "Documents", "aramayo", "GIT", and several "Computational_Genom..." entries. Below that are sections for "iCloud" and "Locations" with "BlackWolf", "Lupus04", and "BlackWolf HD". The main area of the dialog shows a file list with two items:

Name	Size	Kind
Homo_sapiens.GRCh38.pep.all.fa	79.7 MB	sequence
Homo_sapiens.GRCh38.pep.all.fa.gz	14.4 MB	gzip co...archive

An arrow points from the text "Upload file into Kaiser" to the "Downloads" section of the file list. The background of the interface shows a blurred image of a dog and some navigation menus like "Tools", "HPRC", and "History".

BLAST in Galaxy

Upload file into Kaiser

Galaxy https://galaxy-terra.hprc.tamu.edu/kaiser

HPRC Kaiser Galaxy

Tools search tools

HPRC

- Get Data
- Send Data
- Collection Operations
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- Fetch Alignments/Sequences
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- FASTQ Quality Control
- Mapping
- NCBI BLAST+
- Assembly
- Annotation
- FASTA/FASTQ
- Datamash
- EMBOSS
- Multiple Alignments
- Picard
- BED

Download from web or upload from disk

Regular Composite Collection Rule-based

You added 1 file(s) to the queue. Add more files or click 'Start' to proceed.

Name	Size	Type	Genome	Settings	Status
Homo_sapiens.GRC	76 MB	Auto-de...	----- Additional ...	⚙️	0%

Type (set all): Auto-detect Q Genome (set all) ----- Additional ...

COVID-19 related research on Galaxy: [training](#), [tutorials](#), [documents](#)

Uploading data from your computer from Galaxy Project

History search datasets Using 2% Lecture16 (empty) This history is empty. You can load your own data or get data from an external source

BLAST in Galaxy

Upload file into Kaiser

Galaxy https://galaxy-terra.hprc.tamu.edu/kaiser

HPRC Kaiser Galaxy

Tools search tools

HPRC

- Get Data
- Send Data
- Collection Operations
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- Fetch Alignments/Sequences
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- FASTQ Quality Control
- Mapping
- NCBI BLAST+
- Assembly
- Annotation
- FASTA/FASTQ
- Datamash
- EMBOSS
- Multiple Alignments
- Picard
- BED

Download from web or upload from disk

Please wait...1 out of 1 remaining.

Name	Size	Type	Genome	Settings	Status
Homo_sapiens.GRC	76 MB	Auto-de...	----- Additional ...		43%

Type (set all): Auto-detect Genome (set all): ----- Additional ...

Choose local files Choose FTP files Paste/Fetch data Start Select Pause Reset Close

COVID-19 related research on Galaxy: [training](#), [tutorials](#), [documents](#)

Current known issues: Some tools cannot be removed from favorites. Choose your favorites wisely.

Uploading data from your computer from Galaxy Project

History search datasets Lecture16 (empty) This history is empty. You can load your own data or get data from an external source

BLAST in Galaxy

Upload file into Kaiser

The screenshot shows the HPRC Kaiser Galaxy web interface. A modal dialog box is open in the center, titled "Download from web or upload from disk". The dialog contains a table with one row, showing a file named "Homo_sapiens.GRC" which is 76 MB in size, type "Auto-de...", genome "Additional ...", and has a status of "100%". Below the table are buttons for "Choose local files", "Choose FTP files", "Paste/Fetch data", "Start", "Select" (which is highlighted in blue), "Pause", "Reset", and "Close". An arrow points from the text "Upload file into Kaiser" to the "Select" button. The background shows the Galaxy tool menu on the left and a history panel on the right.

Galaxy

https://galaxy-terra.hprc.tamu.edu/kaiser

HPRC Kaiser Galaxy

Analyze Data Workflow Visualize Shared Data Admin Help User

Using 2%

Tools

search tools

HPRC

Get Data

Send Data

Collection Operations

Lift-Over

Text Manipulation

Convert Formats

Filter and Sort

Join, Subtract and Group

Fetch Alignments/Sequences

Operate on Genomic Intervals

Statistics

Graph/Display Data

Phenotype Association

FASTQ Quality Control

Mapping

NCBI BLAST+

Assembly

Annotation

FASTA/FASTQ

Datamash

EMBOSS

Multiple Alignments

Picard

BED

History

search datasets

Lecture16

1 shown

(empty)

1: Homo_sapiens.GRC h38.pep.all.fa

Download from web or upload from disk

Name	Size	Type	Genome	Settings	Status
Homo_sapiens.GRC	76 MB	Auto-de...	Additional ...		100%

Type (set all): Auto-detect Genome (set all): Additional ...

Start Pause Reset Close

COVID-19 related research on Galaxy: [training](#), [tutorials](#), [documents](#)

Current known issues: Some tools cannot be removed from favorites. Choose your favorites wisely.

Uploading data from your computer from Galaxy Project

BLAST in Galaxy

Upload file into Kaiser

Galaxy https://galaxy-terra.hprc.tamu.edu/kaiser

HPRC Kaiser Galaxy

Analyze Data Workflow Visualize Shared Data Admin Help User

Tools search tools

HPRC

Get Data Send Data Collection Operations Lift-Over Text Manipulation Convert Formats Filter and Sort Join, Subtract and Group Fetch Alignments/Sequences Operate on Genomic Intervals Statistics Graph/Display Data Phenotype Association FASTQ Quality Control Mapping NCBI BLAST+ Assembly Annotation FASTA/FASTQ Datamash EMBOSS Multiple Alignments Picard BED

Best Practices for Kaiser Galaxy

- Kaiser Galaxy (docs, slides) is configured for teaching purposes so all users have a file quota of 1TB. How to permanently delete nonessential files.
- FTP uploads are removed from ftp directory 48 hours after uploading so import your ftp files into Galaxy the same day as you upload to ftp
- Only certain tools that support multi-core processing have the Job Resources Parameters option which allow you to select cores, memory and time.
- The default job resource parameters for all tools is 1 core with 2GB memory for 24 hours (24 SUs).
 - Configuring a job to use 28 cores for 1 hour requires 28 SUs. (28 cores for 168 hours = 4704 SUs).
 - Configuring a job to use 54GB memory for 1 hour requires 28 SUs. (54GB memory for 168 hours = 4704 SUs).
 - If a tool you used failed because it needs the Job Resource Parameters option added, contact the HPRC helpdesk.



COVID-19 related research on Galaxy: [training](#), [tutorials](#), [documents](#)

Current known issues: Some tools cannot be removed from favorites. Choose your favorites wisely.

Uploading data from your computer from Galaxy Project

History search datasets

Lecture16 1 shown 76.04 MB

1: Homo_sapiens.GRCh3 8.pep.all.fa

BLAST in Galaxy

Rename file: Hs_EnsRel105_Proteins

The screenshot shows the Galaxy web interface for managing datasets. On the left, a sidebar lists various tools and analysis types. The main area is titled "Edit dataset attributes" for a dataset named "Hs_EnsRel105_Proteins". The dataset is described as an "uploaded fasta file". There are sections for "Info", "Annotation", and "Database/Build". The "Save" button is located at the top right of the edit form. A "History" panel on the right shows a single entry for the dataset.

Galaxy | HPRC Kaiser Galaxy

https://galaxy-terra.hprc.tamu.edu/kaiser/datasets/edit

HPRC Kaiser Galaxy

Tools

- search tools

HPRC

- Get Data
- Send Data
- Collection Operations
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- Fetch Alignments/Sequences
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- FASTQ Quality Control
- Mapping
- NCBI BLAST+
- Assembly
- Annotation
- FASTA/FASTQ
- Datamash
- EMBOSS
- Multiple Alignments
- Picard
- BED

Edit dataset attributes

Name: Hs_EnsRel105_Proteins

Info: uploaded fasta file

Annotation:

Add an annotation or notes to a dataset; annotations are available when a history is viewed.

Database/Build: ----- Additional Species Are Below -----

Save

History

Lecture16

1 shown

76.04 MB

1: Homo_sapiens.GRCh3 8.pep.all.fa

BLAST in Galaxy

Rename file: Hs_EnsRel105_Proteins

The screenshot shows the HPRC Kaiser Galaxy web interface. The main area displays the 'Edit dataset attributes' page for a dataset named 'Hs_EnsRel105_Proteins'. The 'Name' field is highlighted with a red box and a red arrow pointing to it from the top-left. The 'Save' button at the top right is also highlighted with a red box and a red arrow pointing to it from the top-right. In the bottom right corner of the main window, there is a small green box containing the text '1: Hs_EnsRel105_Proteins'. The left sidebar lists various bioinformatics tools under the 'HPRC' section.

Galaxy

https://galaxy-terra.hprc.tamu.edu/kaiser/datasets/edit

HPRC Kaiser Galaxy

Tools

- Get Data
- Send Data
- Collection Operations
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- Fetch Alignments/Sequences
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- FASTQ Quality Control
- Mapping
- NCBI BLAST+
- Assembly
- Annotation
- FASTA/FASTQ
- Datamash
- EMBOSS
- Multiple Alignments
- Picard
- BED

Edit dataset attributes

Attributes updated.

Attributes Convert Datatypes Permissions

Edit attributes

Name

Hs_EnsRel105_Proteins

Info

uploaded fasta file

Annotation

Add an annotation or notes to a dataset; annotations are available when a history is viewed.

Database/Build

----- Additional Species Are Below -----

History

search datasets

Lecture16

1 shown

76.04 MB

1: Hs_EnsRel105_Proteins

BLAST in Galaxy

Preparing BLAST Datasets

- Repeat the same for:
- ftp://ftp.ensembl.org/pub/release-105/fasta/mus_musculus/pep/Mus_musculus.GRCm38.pep.all.fa.gz
- Rename to: Mm_EnsRel105_Proteins

BLAST in Galaxy

Extract the 10 largest proteins from each set

- Run ‘FASTA Tools’ > ‘Compute sequence length’
- Run ‘Filter and Sort’ > ‘Sort data in ascending or descending order’
 - On column 02, Descending, Fast numeric (-n)
- Run ‘Text Manipulation’ > ‘Select first lines from a dataset (head)’
 - Operation: Keep first lines
 - Number of lines: 10
- Run ‘Text Manipulation’ > ‘Convert delimiters to TAB’
 - Convert all Whitespaces
- Run ‘Text Manipulation’ > ‘Cut (reorder) columns from a table’
 - Cut c1
- Run ‘FASTA Tools’ > ‘faSomeRecords Select fasta sequences based on full header content’
 - Input fasta file: Hs_EnsRel90_Proteins
- File containing text to match: Cut (reorder) on data 6
- Rename: ‘Sort Fasta file on data 9’ to ‘Hs_EnsRel105_10LProteins’
- Generate WorkFlow ‘CG2021_WFL_01’
- Run CG2022_WFL_01 on:
 - Mouse (Mm_EnsRel105_Proteins) and
- Rename resulting file as: Mm_EnsRel105_10LProteins

BLAST in Galaxy

Looking for Sequence Homology

Job	Query	Subject	E-Value	Output Format	Advanced Options
1	Hs_EnsRel105_10LProteins	Mm_EnsRel105_10LProteins	1E-05	Pairwise text	None
2	Hs_EnsRel105_10LProteins	Mm_EnsRel105_10LProteins	1E-05	Tabular 12 Columns	None
3	Hs_EnsRel105_10LProteins	Mm_EnsRel105_10LProteins	1E-05	Tabular 12 Columns	MaxSeqs: 02
4	Hs_EnsRel105_10LProteins	Mm_EnsRel105_10LProteins	1E-15	Tabular 12 Columns	MaxSeqs: 02
5	Hs_EnsRel105_10LProteins	Mm_EnsRel105_10LProteins	1E-25	Tabular 12 Columns	MaxSeqs: 02
6	Hs_EnsRel105_10LProteins	Mm_EnsRel105_10LProteins	1E-50	Tabular 12 Columns	MaxSeqs: 02