ITMO3

GALAXY

Initiation Libraries

Le Corguillé
2 GO
GET DATA
2 GO
LIBRARIES
What is a library?

http://abims.sb-roscoff.fr/galaxyproject

galaxy

hg19_chr12.bam

FTP

dataset_74095.dat
• Les programmes de FTP

- WinSCP
- FileZilla
- Cyberduck
- WinSCP

Avoid: Malwares inside
Import d'un fichier > 2 Go
Import d'un fichier > 2 Go
Import d'un fichier > 2 Go
Import d'un fichier > 2 Go
Import d'un fichier > 2 Go
IMPORT DES DONNÉES

> 2 GO - PART II – CÔTÉ GALAXY
# Data Libraries

<table>
<thead>
<tr>
<th>Data library name</th>
<th>Data library description</th>
</tr>
</thead>
<tbody>
<tr>
<td>lecorquille</td>
<td>Dataset for RNA-seq de-novo, re-ingeniered - pporicard</td>
</tr>
<tr>
<td>RNA-seq de-novo</td>
<td>Dataset for RNA-seq de-novo, re-ingeniered - pporicard</td>
</tr>
<tr>
<td>RNA-seq reference</td>
<td>Dataset for RNA-seq with reference genome - acormier</td>
</tr>
</tbody>
</table>

**ONE USER = ONE LIBRARY**
Create a new folder

**Name:**

RNASeq

**Description:**

Create
The new folder named 'RNASEq' has been added to the data library.

<table>
<thead>
<tr>
<th>Name</th>
<th>Message</th>
<th>Data type</th>
<th>Date uploaded</th>
<th>File size</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNASEq</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Add datasets
- Add sub-folder
- Edit information
- Move this folder
- Use template
- Delete this folder
Upload files to a data library

Upload a directory of files

Upload option:
- Upload directory of files

Choose upload option (file, directory, filesystem paths, current history).

File Format:
- Auto-detect

Server Directory:
- masseq-training

Upload all files in a sub-directory of /project/sb/galaxyimport/user/lecorguille@sb-roscoff.fr on the Galaxy server.

Copy data into Galaxy?
- Copy files into Galaxy

Copy files into Galaxy

Normally data uploaded with this tool is copied into Galaxy's configured "file_path" location where Galaxy has a form of control over the data files. However, this may not be desired (especially for large NGS datasets), so using the option labeled "Link to files without copying into Galaxy" will force Galaxy to always read the data from its original path. Any symlinks encountered in the uploaded directory will be dereferenced once. That is, Galaxy will point directly to the file that is linked, but no other files in the subdirectory will be dereferenced.

Convert spaces to tabs:
- Yes

Use this option if you are entering intervals by hand.

Genome:
- unspecified (?)

Message:

This information will be displayed in the "Message" column for this dataset in the data library browser.

Restrict dataset access to specific roles:
- lecorguille@sb-roscoff.fr

Multi-select list - hold the appropriate key while clicking to select multiple roles. More restrictions can be applied after the upload is complete. Selecting no roles makes a dataset public.

Link to files without copying
Copy files into Galaxy
### Data Library “lecorguille”

- Added 4 datasets to the folder ‘RNASeq’ (each is selected).

<table>
<thead>
<tr>
<th>Name</th>
<th>Message</th>
<th>Data type</th>
<th>Date uploaded</th>
<th>File size</th>
</tr>
</thead>
<tbody>
<tr>
<td>BlueLight.sample.paired.1.cleaned.fastq</td>
<td>This job is running</td>
<td>auto</td>
<td>Mon Oct 6 09:02:32 2014 (UTC)</td>
<td>0 bytes</td>
</tr>
<tr>
<td>BlueLight.sample.paired.2.cleaned.fastq</td>
<td>This job is running</td>
<td>auto</td>
<td>Mon Oct 6 09:02:32 2014 (UTC)</td>
<td>0 bytes</td>
</tr>
<tr>
<td>Dark.sample.paired.1.cleaned.fastq</td>
<td>This job is running</td>
<td>auto</td>
<td>Mon Oct 6 09:02:33 2014 (UTC)</td>
<td>0 bytes</td>
</tr>
<tr>
<td>Dark.sample.paired.2.cleaned.fastq</td>
<td>This job is running</td>
<td>auto</td>
<td>Mon Oct 6 09:02:33 2014 (UTC)</td>
<td>0 bytes</td>
</tr>
</tbody>
</table>

**TIP:** You can download individual library datasets by selecting “Download this dataset” from the context menu (triangle) next to each dataset’s name.

**TIP:** Several compression options are available for downloading multiple library datasets simultaneously:
- gzip: Recommended for fast network connections
- bzip2: Recommended for slower network connections (smaller size but takes longer to compress)
- zip: Not recommended but is provided as an option for those who cannot open the above formats
Data Libraries Beta Test. This is work in progress. Please give your feedback, comments & ideas via email and Trello.

<table>
<thead>
<tr>
<th>name</th>
<th>size</th>
<th>date (UTC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BlueLight sample paired 1 cleaned fastq</td>
<td>fastq</td>
<td>0.1GB</td>
</tr>
<tr>
<td>BlueLight sample paired 2 cleaned fastq</td>
<td>fastq</td>
<td>0.1GB</td>
</tr>
<tr>
<td>Dark sample paired 1 cleaned fastq</td>
<td>fastq</td>
<td>95.3MB</td>
</tr>
<tr>
<td>Dark sample paired 2 cleaned fastq</td>
<td>fastq</td>
<td>93.9MB</td>
</tr>
</tbody>
</table>