



Biogenouest  
BIOGÉNOUWEST



ifb  
INSTITUT FRANÇAIS  
DE BIOINFORMATIQUE



EMBRC  
FRANCE  
CENTRE NATIONAL DE RESSOURCES BIOLOGIQUES MARINES



OCEANOMICS



IDEALG  
seaweed for the future

# Abims<sup>4</sup>

15/05/2014



Galaxy

## Initiation

Le Corguillé – Monsoor

V1.01

UPMC  
SORBONNE UNIVERSITÉS



# INTRODUCTION / PROBLÉMATIQUE

- Setup TP
  - <http://galaxy.sb-roscoff.fr>

- Compte

- login@sb-roscoff.fr
- \*\*\*\*\*

- Support de cours :

<http://application.sb-roscoff.fr/download/fr2424/abims/lecorguille/cours/galaxy-initiation.pdf>

Sélectionner votre niveau :

Level 1



“Je veux connaître l'expression de mes gènes”

Level 2



“Je veux faire un mapping et un comptage de mes reads sur mon génome de référence”

## Level 3



“Je veux lancer les outils tophat2 et cufflinks  
J'ai des fastq et mon génome au format fasta et gtf”

Level 4



“Je veux un espace projet avec 1To car je pense faire du ssh pour lancer tophat2 sur 8 procs via qsub et soumettre le fichier bam à mon génome avec cufflinks Sinon, pas besoin de vous :P”



Level 5



“J'ai un tas d'outils sympas !  
Mais je suis le seul à pouvoir les lancer.

Des commentaires ?”

# Pourquoi ?

```
login@sbr4-1042:~$ ssh -Y login@bioinfo.sb-roscoff.fr
[...]
[login@n0 ~]$ cd projet
[login@n0 login]$ cd 13-07-29-panda/tmp/mapping
[login@n0 mapping]$ cat tophat.qsub
#!/bin/bash
#$ -S /bin/bash
#$ -M login@sb-roscoff.fr
#$ -m bea
#$ -V
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err

tophat2 panda_v121029 ../input/I11R1-1.fq ../input/I11R1-2.fq
-GTF ../input/panda_v121029.gtf --b2-sensitive -r 100
-num-threads 8

[login@n0 mapping]$ qsub -q long.q -pe thread 8 tophat.qsub
Your job 5338969 ("tophat.qsub") has been submitted
[login@n0 mapping]$ ls
accepted_hits.bam    junctions.bed        qsub.err    unmapped.bam
deletions.bed       logs                 qsub.out
insertions.bed      prep_reads.info     tmp
[login@n0 mapping]$ cd ..
[login@n0 mapping]$ mkdir cufflinks
```

```
login@sbr4-1042:~$ ssh -Y login@bioinfo.sb-roscoff.fr
[...]
```

```
[login@n0 ~]$ cd projet
[login@n0 login]$
[login@n0 mapping]$
#!/bin/bash
#$ -S /bin/bash
#$ -M login@sb-roscoff.fr
#$ -m bea
#$ -V
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err
```



```
tophat2 panda_v1.2.1 -R1-2.fq
-GTF ../input/panda_v1.2.1.gtf
-num-threads 8
```

R1-2.fq

```
[login@n0 mapping]$
Your job 5338969
[login@n0 mapping]$
accepted_hits.bam
deletions.bed
insertions.bed
[login@n0 mapping]$ cd ..
[login@n0 mapping]$ mkdir cufflinks
```

qsub

1.bam

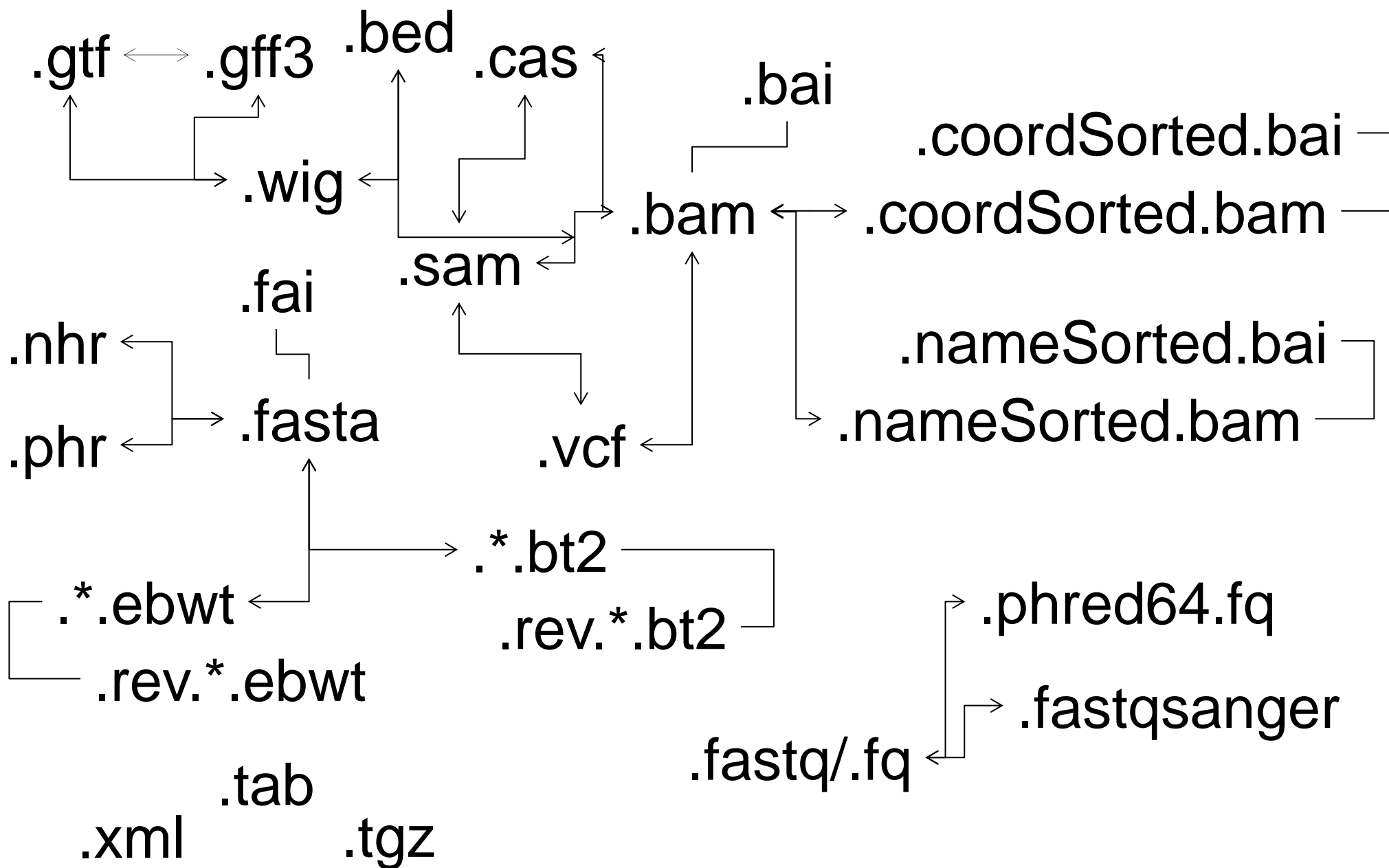
**NOOOOOOOO!**

```
prep_reads.info tmp
```

# Introduction

.gtf .gff3 .bed .cas .bai  
.wig .bam .coordSorted.bai  
.fai .sam .coordSorted.bam  
.nhr .nameSorted.bai  
.phr .fasta .vcf .nameSorted.bam  
.\*.ebwt \*.bt2 .phred64.fq  
.rev.\*.ebwt .rev.\*.bt2 .fastqsanger  
.tab .fastq/.fq  
.xml .tgz

# Introduction



- Les outils avec interface graphique click bouton sous windows
  - + très ergonomique
  - trop ergonomique → manque de souplesse
  - faut pas rêver ! Vous avez déjà vu un thésard trouver le temps pour faire de beaux boutons verts ?
- Les outils en ligne sur Internet
  - + très ergonomique
  - trop ergonomique → manque de souplesse
  - une infime part des outils disponibles
  - réparti un peu partout sur les différents sites universitaires
  - souvent limité en terme de taille de soumission
  - il ne faut pas être parano

- Les outils en ligne de commande
  - + représente la quasi majorité des outils scientifiques
  - + très complet en terme de paramètre
  - + peuvent tourner sur des clusters de calcul
  - + les g33ks adorent car automatisable, workflowsable, ...
  - nécessite un minimum de connaissance en linux
  - manque cruellement d'ergonomie



# INTRODUCTION / GALAXY

Tools

search tools

**Get Data**

- Upload File from your computer

ABiMS WORKFLOWS

- Workflow RNA-seq de novo by ABiMS
- Workflow RNA-seq with reference by ABiMS
- Workflow 4 Metabolomics

ABiMS TOOLS

- Primer
- RNASeq
- InterEsil
- Statistics
- Utils
- Phylogenetics
- Debug

COMMON TOOLS

- Text Manipulation
- FASTA manipulation
- Join, Subtract and Group
- Filter and Sort
- NCBI BLAST+**
  - NCBI BLAST+ blastn Search nucleotide database with nucleotide query sequence(s)
  - NCBI BLAST+ blastp Search protein database with protein query sequence(s)
  - NCBI BLAST+ blastx Search protein database with translated nucleotide query sequence(s)
  - NCBI BLAST+ tblastn Search translated nucleotide database

### NCBI BLAST+ blastx (version 0.0.17)

**Nucleotide query sequence(s):**  
1: human\_protein.fas

**Subject database/sequences:**  
FASTA file from your history (see warning note below)

**Protein FASTA file to use as database:**  
1: human\_protein.fas

**Query genetic code:**  
1. Standard

**Set expectation value cutoff:**  
0.001

**Output format:**  
Tabular (extended 24 columns)

**Advanced Options:**  
Hide Advanced Options

**Execute**

**Note.** Database searches may take a substantial amount of time. For large input datasets it is advisable to allow overnight processing.

**What it does**  
Search a *protein database* using a *translated nucleotide query*, using the NCBI BLAST+ blastx command line tool.

**Note** You can also search against a FASTA file of subject protein sequences. This is *not* advised because it is slower (only one CPU is used), but more importantly gives e-values for pairwise searches (very small e-values which will look overly significant). In most cases you should instead turn the other FASTA file into a database first using *makeblastdb* and search against that.

**Output format**  
Because Galaxy focuses on processing tabular data, the default output of this tool is tabular. The standard BLAST+ tabular output contains 12 columns:

Column	NCBI name	Description
--------	-----------	-------------

**History**

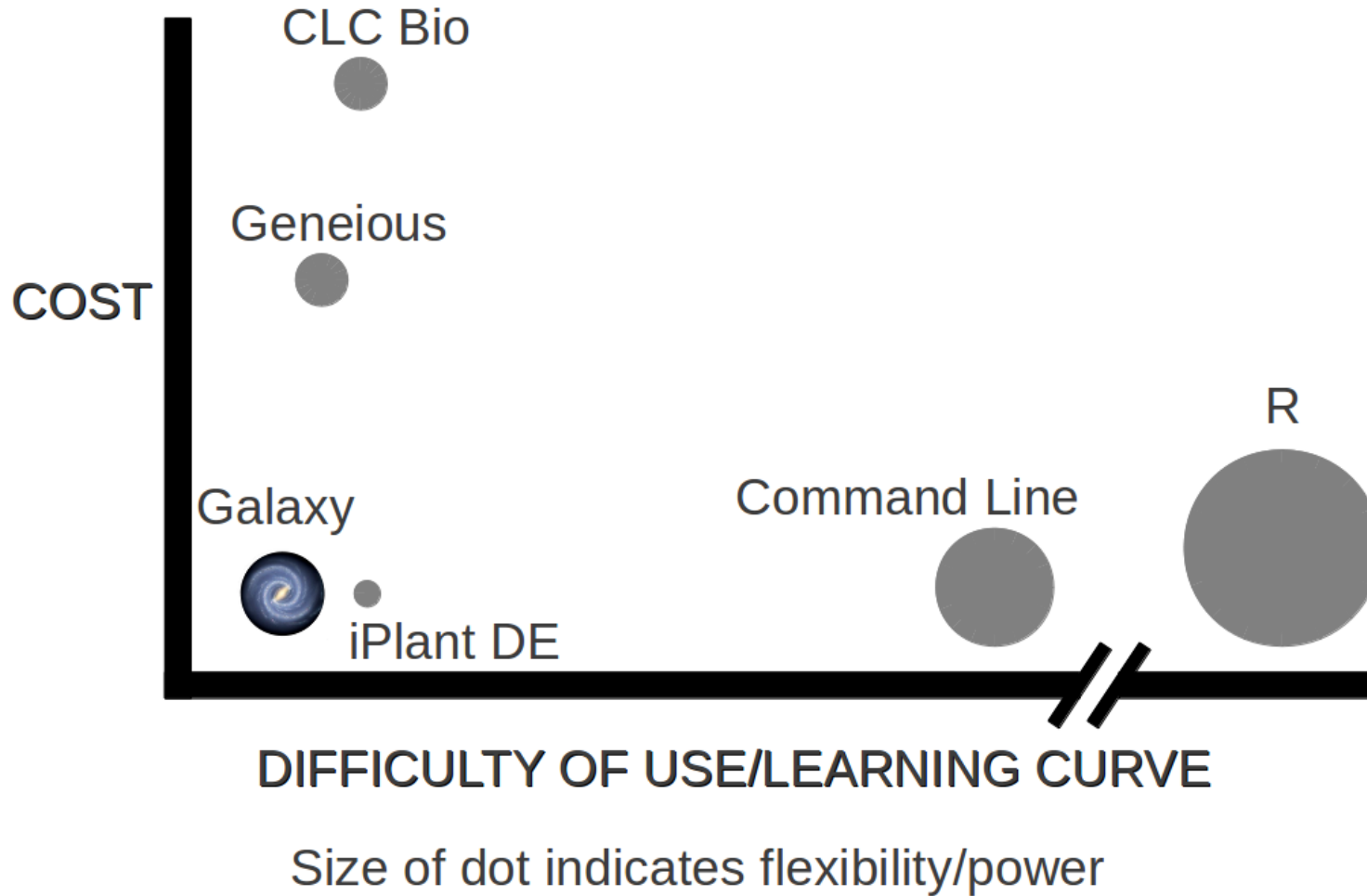
- Human protein study  
5.3 MB
- 2: chr22\_check.gff3
- 1: human\_protein.fas

- Galaxy c'est ...
  - Plus besoin de lancer une ligne de commande dans un terminal
  - Plus besoin de connaître la programmation ou le scripting
  - Des jobs soumis de manière transparente sur un cluster de calcul
  - Un gestionnaire d'historique et de données sécurisées
  - Un système de partage de données ou de protocoles
  - Des boîtes à outils dans plusieurs domaines de la bioinformatique
    - NGS
    - Chimie
    - Metabolomique
    - Etc ...
    - Statistique
    - Analyse d'image



# Introduction

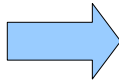
## RNA-Seq Analysis Tools



- Pourquoi Galaxy ?
  - Accessibilité
    - Accès à des outils de bioinformatique sans connaissance en informatique
    - Ergonomie à géométrie variable
    - Modularité
  - Reproductibilité
    - Traçabilité des paramètres
  - Transparence
    - Partage des données et protocoles

# Introduction

## MR. GEEK



```
[login@n0 ~]$ cdprojct
[login@n0 login]$ cd 13-07-29-panda/tmp/mapping
[login@n0 mapping]$ cat tophat.qsub
#!/bin/bash
#$ -S /bin/bash
#$ -M login@sb-roscoff.fr
#$ -m bea
#$ -v
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err

tophat2 panda_v121029 ../input/I11R1-1.fq ../input/I11R1-2.fq
-GTF ../input/panda_v121029.gtf --b2-sensitive -r 100
-num-threads 8

[login@n0 mapping]$ qsub -q long.q -pe thread 8 tophat.qsub
Your job 5338969 ("tophat.qsub") has been submitted
```

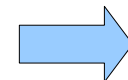


# Introduction

```
[login@n0 ~]$ cdprojet
[login@n0 login]$ cd 13-07-29-panda/tmp/mapping
[login@n0 mapping]$ cat tophat.qsub
#!/bin/bash
#$ -S /bin/bash
#$ -M login@sb-roscoff.fr
#$ -m bea
#$ -v
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err

tophat2 panda_v121029 ../input/I11R1-1.fq ../input/I11R1-2.fq
-GTF ../input/panda_v121029.gtf --b2-sensitive -r 100
-num-threads 8

[login@n0 mapping]$ qsub -q long.q -pe thread 8 tophat.qsub
Your job 5338969 ("tophat.qsub") has been submitted
```



**MR. HAPPY**

*by Roger Hargreaves*



# Introduction

```
[lecorguille@n0 ~]$ e-PCR --help
e-PCR: invalid option -- -
usage: [-hV] [posix-options] stsfile [fasta ...]
[compat-options]
where posix-options are:
    -m ##      Margin (default 50)
    -w ##      Wordsize (default 7)
    -n ##      Max mismatches allowed (default 0)
    -g ##      Max indels allowed (default 0)
    -f ##      Use ## discontinuous words, slow if
                ##>1
    -o ##      Set output file
    -t ##      Set output format:
                1 - classic, range (pos1..pos2)
                2 - classic, midpoint
                3 - tabular
                4 - tabular with alignment in
                    (slow)
    -d##-##    Set default size range
                (default 100-350)
    -p +-      Turn hits postprocess on/off
    -v ##      Verbosity flags
    -a a|f     Use presize alignmens (only if
                gaps>0), slow
                a - Always or f - as Fallback
    -x +-      Use 5'-end lowercase masking of
                primers (default -)
    -u +-      Uppercase all primers (default -)

[...]
```

Galaxy / ABiMS

e-PCR (version 1.0.0)

**STS file:**  
  
 format : tabular

**Fasta file:**  
  
 format : fasta

**Wordsize (W):**  
  
 Set word size for primers hash (nucleotide positions). Longer word size decreases hash collision rate, but increases memory usage. Also no mismatches are allowed within word size near 'inner' boundary of primers unless one uses discontinuous words, and no gaps are ever allowed in that region.

**Use ## discontinuous words (F):**  
  
 Set discontinuous word count for primers hash (1 means 'use contiguous words'). Discontinuous words increase number of hash tables and decrease 'effective' word size (thus increasing hash collision rate), so make search significantly slower, but increase sensitivity by allowing mismatches within word size. Reasonable values are 1 (contiguous words) and 3.

**Margin (M):**  
  
 Set maximal allowed deviation of hit product size from expected STS size.

**Set default sts lower size (D):**  
  
 Set ddefault STS size range - values used for STSs that have no size associated in file.

**Set default sts higher size (D):**  
  
 Set ddefault STS size range - values used for STSs that have no size associated in file.

**Max mismatches allowed (N):**  
  
 Set maximal number of mismatches allowed in primer-to-sequence alignment (per primer!).

**Max indels allowed (G):**  
  
 Set maximal number of gaps allowed in primer-to-sequence alignment (per primer!).

**Set output format (T):**  
  
 Output formats



Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Help User Using 41%

Tools

search tools

**Get Data**

- Upload File from your computer

ABiMS WORKFLOWS

- Workflow RNA-seq de novo by ABiMS
- Workflow RNA-seq with reference by ABiMS
- Workflow 4 Metabolomics

ABiMS TOOLS

- Primer
- RNASeq
- InterEsil
- Statistics
- Utils
- Phylogenetics
- Debug

COMMON TOOLS

- Text Manipulation
- FASTA manipulation
- Join, Subtract and Group
- Filter and Sort
- NCBI BLAST+

- NCBI BLAST+ `blastn` Search nucleotide database with nucleotide query sequence(s)
- NCBI BLAST+ `blastp` Search protein database with protein query sequence(s)
- NCBI BLAST+ `blastx` Search protein database with translated nucleotide query sequence(s)
- NCBI BLAST+ `tblastn` Search translated nucleotide database

### NCBI BLAST+ blastx (version 0.0.17)

**Nucleotide query sequence(s):**  
1: human\_protein.fas

**Subject database/sequences:**  
FASTA file from your history (see warning note below)

**Protein FASTA file to use as database:**  
1: human\_protein.fas

**Query genetic code:**  
1. Standard

**Set expectation value cutoff:**  
0.001

**Output format:**  
Tabular (extended 24 columns)

**Advanced Options:**  
Hide Advanced Options

**Execute**

**Note.** Database searches may take a substantial amount of time. For large input datasets it is advisable to allow overnight processing.

**What it does**

Search a *protein database* using a *translated nucleotide query*, using the NCBI BLAST+ `blastx` command line tool.

**Note** You can also search against a FASTA file of subject protein sequences. This is *not* advised because it is slower (only one CPU is used), but more importantly gives e-values for pairwise searches (very small e-values which will look overly significant). In most cases you should instead turn the other FASTA file into a database first using `makeblastdb` and search against that.

**Output format**

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

Column	NCBI name	Description
--------	-----------	-------------

History

- Human protein study  
5.3 MB
- 2: chr22\_check.gff3
- 1: human\_protein.fas

25 /

## le menu

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Help User Using 41%

Tools

search tools

Get Data

- Upload File from your computer

ABiMS WORKFLOWS

- Workflow RNA-seq de novo by ABiMS
- Workflow RNA-seq with reference by ABiMS
- Workflow 4 Metabolomics

ABiMS TOOLS

- Primer
- RNASeq
- InterEsil
- Statistics
- Utils
- Phylogenetics
- Debug

COMMON TOOLS

- Text Manipulation
- FASTA manipulation
- Join, Subtract and Group
- Filter and Sort
- NCBI BLAST+

- NCBI BLAST+ `blastn` Search nucleotide database with nucleotide query sequence(s)
- NCBI BLAST+ `blastp` Search protein database with protein query sequence(s)
- NCBI BLAST+ `blastx` Search protein database with translated nucleotide query sequence(s)
- NCBI BLAST+ `tblastn` Search translated nucleotide database

NCBI BLAST+ blastx (version 0.0.17)

Nucleotide query sequence(s):  
1: human\_protein.fas

Subject database/sequences:  
FASTA file from your history (see warning note below)

Protein FASTA file to use as database:  
1: human\_protein.fas

Query genetic code:  
1. Standard

Set expectation value cutoff:  
0.001

Output format:  
Tabular (extended 24 columns)

Advanced Options:  
Hide Advanced Options

Execute

**Note.** Database searches may take a substantial amount of time. For large input datasets it is advisable to allow overnight processing.

**What it does**

Search a *protein database* using a *translated nucleotide query*, using the NCBI BLAST+ `blastx` command line tool.

**Note** You can also search against a FASTA file of subject protein sequences. This is *not* advised because it is slower (only one CPU is used), but more importantly gives e-values for pairwise searches (very small e-values which will look overly significant). In most cases you should instead turn the other FASTA file into a database first using `makeblastdb` and search against that.

**Output format**

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

Column	NCBI name	Description
--------	-----------	-------------

History

- Human protein study  
5.3 MB
- 2: chr22\_check.gff3
- 1: human\_protein.fas

26 /

## la liste des outils

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Help User Using 41%

**Tools**

search tools

**Get Data**

- Upload File from your computer

**ABiMS WORKFLOWS**

- Workflow RNA-seq de novo by ABiMS
- Workflow RNA-seq with reference by ABiMS
- Workflow 4 Metabolomics

**ABiMS TOOLS**

- Primer
- RNASeq
- InterEsil
- Statistics
- Utils
- Phylogenetics
- Debug

**COMMON TOOLS**

- Text Manipulation
- FASTA manipulation
- Join, Subtract and Group
- Filter and Sort
- NCBI BLAST+**
  - NCBI BLAST+ `blastn` Search nucleotide database with nucleotide query sequence(s)
  - NCBI BLAST+ `blastp` Search protein database with protein query sequence(s)
  - NCBI BLAST+ `blastx` Search protein database with translated nucleotide query sequence(s)
  - NCBI BLAST+ `tblastn` Search translated nucleotide database

**NCBI BLAST+ blastx (version 0.0.17)**

**Nucleotide query sequence(s):**  
1: human\_protein.fas

**Subject database/sequences:**  
FASTA file from your history (see warning note below)

**Protein FASTA file to use as database:**  
1: human\_protein.fas

**Query genetic code:**  
1. Standard

**Set expectation value cutoff:**  
0.001

**Output format:**  
Tabular (extended 24 columns)

**Advanced Options:**  
Hide Advanced Options

Execute

**Note.** Database searches may take a substantial amount of time. For large input datasets it is advisable to allow overnight processing.

**What it does**

Search a *protein database* using a *translated nucleotide query*, using the NCBI BLAST+ `blastx` command line tool.

**Note** You can also search against a FASTA file of subject protein sequences. This is *not* advised because it is slower (only one CPU is used), but more importantly gives e-values for pairwise searches (very small e-values which will look overly significant). In most cases you should instead turn the other FASTA file into a database first using `makeblastdb` and search against that.

**Output format**

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

Column	NCBI name	Description
--------	-----------	-------------

**History**

- Human protein study  
5.3 MB
- 2: chr22\_check.gff3
- 1: human\_protein.fas

27 /

## formulaire / visualisation / information diverse

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Help User Using 41%

Tools

search tools

Get Data

- Upload File from your computer

ABiMS WORKFLOWS

- Workflow RNA-seq de novo by ABiMS
- Workflow RNA-seq with reference by ABiMS
- Workflow 4 Metabolomics

ABiMS TOOLS

- Primer
- RNASeq
- InterEsil
- Statistics
- Utils
- Phylogenetics
- Debug

COMMON TOOLS

- Text Manipulation
- FASTA manipulation
- Join, Subtract and Group
- Filter and Sort
- NCBI BLAST+

- NCBI BLAST+ `blastn` Search nucleotide database with nucleotide query sequence(s)
- NCBI BLAST+ `blastp` Search protein database with protein query sequence(s)
- NCBI BLAST+ `blastx` Search protein database with translated nucleotide query sequence(s)
- NCBI BLAST+ `tblastn` Search translated nucleotide database

### NCBI BLAST+ blastx (version 0.0.17)

**Nucleotide query sequence(s):**  
1: human\_protein.fas

**Subject database/sequences:**  
FASTA file from your history (see warning note below)

**Protein FASTA file to use as database:**  
1: human\_protein.fas

**Query genetic code:**  
1. Standard

**Set expectation value cutoff:**  
0.001

**Output format:**  
Tabular (extended 24 columns)

**Advanced Options:**  
Hide Advanced Options

**Execute**

**Note.** Database searches may take a substantial amount of time. For large input datasets it is advisable to allow overnight processing.

**What it does**

Search a *protein database* using a *translated nucleotide query*, using the NCBI BLAST+ `blastx` command line tool.

**Note** You can also search against a FASTA file of subject protein sequences. This is *not* advised because it is slower (only one CPU is used), but more importantly gives e-values for pairwise searches (very small e-values which will look overly significant). In most cases you should instead turn the other FASTA file into a database first using `makeblastdb` and search against that.

**Output format**

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

Column	NCBI name	Description
--------	-----------	-------------

History

- Human protein study  
5.3 MB
- 2: chr22\_check.gff3
- 1: human\_protein.fas

28 /

## l'historique

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Help User Using 41%

Tools

search tools

Get Data

- Upload File from your computer

ABiMS WORKFLOWS

- Workflow RNA-seq de novo by ABiMS
- Workflow RNA-seq with reference by ABiMS
- Workflow 4 Metabolomics

ABiMS TOOLS

- Primer
- RNASeq
- InterEsil
- Statistics
- Utils
- Phylogenetics
- Debug

COMMON TOOLS

- Text Manipulation
- FASTA manipulation
- Join, Subtract and Group
- Filter and Sort
- NCBI BLAST+

- NCBI BLAST+ `blastn` Search nucleotide database with nucleotide query sequence(s)
- NCBI BLAST+ `blastp` Search protein database with protein query sequence(s)
- NCBI BLAST+ `blastx` Search protein database with translated nucleotide query sequence(s)
- NCBI BLAST+ `tblastn` Search translated nucleotide database

NCBI BLAST+ blastx (version 0.0.17)

Nucleotide query sequence(s):

1: human\_protein.fas

Subject database/sequences:

FASTA file from your history (see warning note below)

Protein FASTA file to use as database:

1: human\_protein.fas

Query genetic code:

1. Standard

Set expectation value cutoff:

0.001

Output format:

Tabular (extended 24 columns)

Advanced Options:

Hide Advanced Options

Execute

**Note.** Database searches may take a substantial amount of time. For large input datasets it is advisable to allow overnight processing.

**What it does**

Search a *protein database* using a *translated nucleotide query*, using the NCBI BLAST+ `blastx` command line tool.

**Note** You can also search against a FASTA file of subject protein sequences. This is *not* advised because it is slower (only one CPU is used), but more importantly gives e-values for pairwise searches (very small e-values which will look overly significant). In most cases you should instead turn the other FASTA file into a database first using `makeblastdb` and search against that.

**Output format**

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

Column	NCBI name	Description
--------	-----------	-------------

History

- Human protein study  
5.3 MB
- 2: chr22\_check.gff3
- 1: human\_protein.fas

29 /

# IMPORT DES DONNÉES

# IMPORT DES DONNÉES

< 2 GO

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Help User Using 42%

**Tools**

search tools

**Get Data** 1

- Upload File from your computer

ABiMS WORKFLOWS

- Workflow RNA-seq de novo by ABiMS
- Workflow RNA-seq with reference by ABiMS
- Workflow 4 Metabolomics

ABiMS TOOLS

- Primer
- RNASeq
- InterEsil
- Statistics
- Utils
- Phylogenetics
- Debug

COMMON TOOLS

- Text Manipulation
- FASTA manipulation
- Join, Subtract and Group
- Filter and Sort
- NCBI BLAST+
- NGS: QC and manipulation
- NGS: RNA Analysis
- NGS: Mapping
- NGS: Picard (beta)
- NGS: SAM Tools
- SVDetect
- VarScan

Workflows

- All workflows

### Upload File (version 1.1.3)

**File Format:**

Auto-detect ▼

Which format? See help below

**File:** 2

TIP: Due to browser limitations, uploading files larger than 2GB is guaranteed to fail. To upload large files, use the URL method (below) or FTP (if enabled by the site administrator).

**URL/Text:**

Here you may specify a list of URLs (one per line) or paste the contents of a file.

**Convert spaces to tabs:**

Yes

Use this option if you are entering intervals by hand.

**Genome:**

unspecified (?) ▼

**Execute** 3

**Auto-detect**

The system will attempt to detect Axt, Fasta, Fastqsolexa, Gff, Gff3, Html, Lav, Maf, Tabular, Wiggle, Bed and Interval (Bed with headers) formats. If your file is not detected properly as one of the known formats, it most likely means that it has some format problems (e.g., different number of columns on different rows). You can still coerce the system to set your data to the format you think it should be. You can also upload compressed files, which will automatically be decompressed.

---

**Ab1**

A binary sequence file in 'ab1' format with a '.ab1' file extension. You must manually select this 'File Format' when uploading the file.

---

**Axt**

blastz pairwise alignment format. Each alignment block in an axt file contains three lines: a summary line and 2 sequence lines. Blocks

**History**

Unnamed history

0 bytes

Your history is empty. Click 'Get Data' on the left pane to start



# Import par copier/coller dans la zone de texte

Il est aussi possible d'y recopier une url ftp d'un fichier en ligne zippé (zip, tgz, gz)

ex : ftp://ftp.ncbi.nih.gov/blast/db/FASTA/mito.aa.gz

Mais attention à la taille !

**Galaxy / ABiMS** Analyze Data Workflow Shared Data Visualization Help User Using 42%

**Tools**  
 search tools  
**Get Data** 1  
 Upload File from your computer

**ABiMS WORKFLOWS**  
 Workflow RNA-seq de novo by ABiMS  
 Workflow RNA-seq with reference by ABiMS  
 Workflow 4 Metabolomics

**ABiMS TOOLS**  
 Primer  
 RNASeq  
 InterEsil  
 Statistics  
 Utils  
 Phylogenetics  
 Debug

**COMMON TOOLS**  
 Text Manipulation  
 FASTA manipulation  
 Join, Subtract and Group  
 Filter and Sort  
 NCBI BLAST+  
 NGS: QC and manipulation  
 NGS: RNA Analysis  
 NGS: Mapping  
 NGS: Picard (beta)  
 NGS: SAM Tools  
 SVDetect  
 VarScan

**Workflows**  
 All workflows

**Upload File (version 1.1.3)**

**File Format:**  
 Auto-detect  
 Which format? See help below

**File:**  
 Browse...

TIP: Due to browser limitations, uploading files larger than 2GB is guaranteed to fail. To upload large files, use the URL method (below) or FTP (if enabled by the site administrator).

**URL/Text:** 2

```
>gjj225735562:5001-89734 Homo sapiens insulin-like growth factor 1 (somatomedin C) (IGF1),
RefSeqGene on chromosome 12
TTTTGTAGATAAATGTGAGGATTTTCTCTAAATCCCTCTTCTGTTTGCTAAATCTCACTGTCACCTGCTAA
ATTCAGAGCAGATAGAGCCTGCGCAATGGAATAAAGTCTCAAATGAAATGTGACATTGCTCTCAACA
TCTCCCATCTCTCTGGATTTCTTTTGTCTCATTATTCCTGCTAACCAATTCATTTTCAGACTTTGTACT
TCAGAAGCAATGGGAAAAATCAGCAGCTTCCAACCCAATTTAAGTGCTGCTTTTGTGATTTCTTGA
AGGTAATATTTCTTACTCTTTGAAGTCATTGGGGAATTCATTTAAATTGTGACTGTTTGCTTCTGCC
TAGAACTGTTCTTCACTTTAAAATTTTCATTGTTCCGGAACCGAGAGTTATTTATAAATTGCTGAATATG
CAATTCGTGGAATCTGAAAAATAGCTCGGGGAGATGGATGCATTTGCACAGATATCTGTATGAGTAGAA
ACTATTGCAAGGTACTTATGCTAAATCCTCCACTTCTGCAGGGCTCCGTGGTGTCATTACAGAAGATTC
CTTTAAATCCTCTCTATGGCTAAGGGCTATAGAGCATGGATATGAACTTGGGGATTTTTTTTTCTTTTG
CAGGTGCAGATGTTTTTTTTAAGACCATGTTCTTTTGCATGTGTGTATGTGTCCTCTGTGTGTATGTGT
```

Here you may specify a list of URLs (one per line) or paste the contents of a file.

**Convert spaces to tabs:**  
 Yes  
 Use this option if you are entering intervals by hand.

**Genome:**  
 unspecified (?)

**Execute** 4

**Auto-detect**  
 The system will attempt to detect Axt, Fasta, Fastqsolexa, Gff, Gff3, Html, Lav, Maf, Tabular, Wiggle, Bed and Interval (Bed with headers) formats. If your file is not detected properly as one of the known formats, it most likely means that it has some format problems (e.g., different number of columns on different rows). You can still coerce the system to set your data to the format you think it should be. You can also upload compressed files, which will automatically be decompressed.

**History**  
 Unnamed history  
 0 bytes  
 Your history is empty. Click 'Get Data' on the left pane to start

Ab1

- tabular → gff
- fastq → fastqsanger
- xml → blastxml

**Galaxy / ABiMS** Analyze Data Workflow Shared Data Visualization Help User Using 42%

Tools

search tools

**Get Data** 1

- Upload File from your computer

ABiMS WORKFLOWS

- Workflow RNA-seq de novo by ABiMS
- Workflow RNA-seq with reference by ABiMS
- Workflow 4 Metabolomics

ABiMS TOOLS

- Primer
- RNASeq
- InterEsil
- Statistics
- Utils
- Phylogenetics
- Debug

COMMON TOOLS

- Text Manipulation
- FASTA manipulation
- Join, Subtract and Group
- Filter and Sort
- NCBI BLAST+
- NGS: QC and manipulation
- NGS: RNA Analysis
- NGS: Mapping
- NGS: Picard (beta)
- NGS: SAM Tools
- SVDetect
- VarScan

Workflows

- All workflows

### Upload File (version 1.1.3)

**File Format:**

Auto-detect 3

ta

csfasta 2

fasta

rdata

tabular

taxonomy

Browse...

Here you may specify a list of URLs (one per line) or paste the contents of a file.

**Convert spaces to tabs:**

Yes

Use this option if you are entering intervals by hand.

**Genome:**

unspecified (?)

**Execute** 4

**Auto-detect**

The system will attempt to detect Axt, Fasta, Fastqsolexa, Gff, Gff3, Html, Lav, Maf, Tabular, Wiggle, Bed and Interval (Bed with headers) formats. If your file is not detected properly as one of the known formats, it most likely means that it has some format problems (e.g., different number of columns on different rows). You can still coerce the system to set your data to the format you think it should be. You can also upload compressed files, which will automatically be decompressed.

**Ab1**

A binary sequence file in 'ab1' format with a '.ab1' file extension. You must manually select this 'File Format' when uploading the file.

**Axt**

blastz pairwise alignment format. Each alignment block in an axt file contains three lines: a summary line and 2 sequence lines. Blocks

**History**

Unnamed history

0 bytes

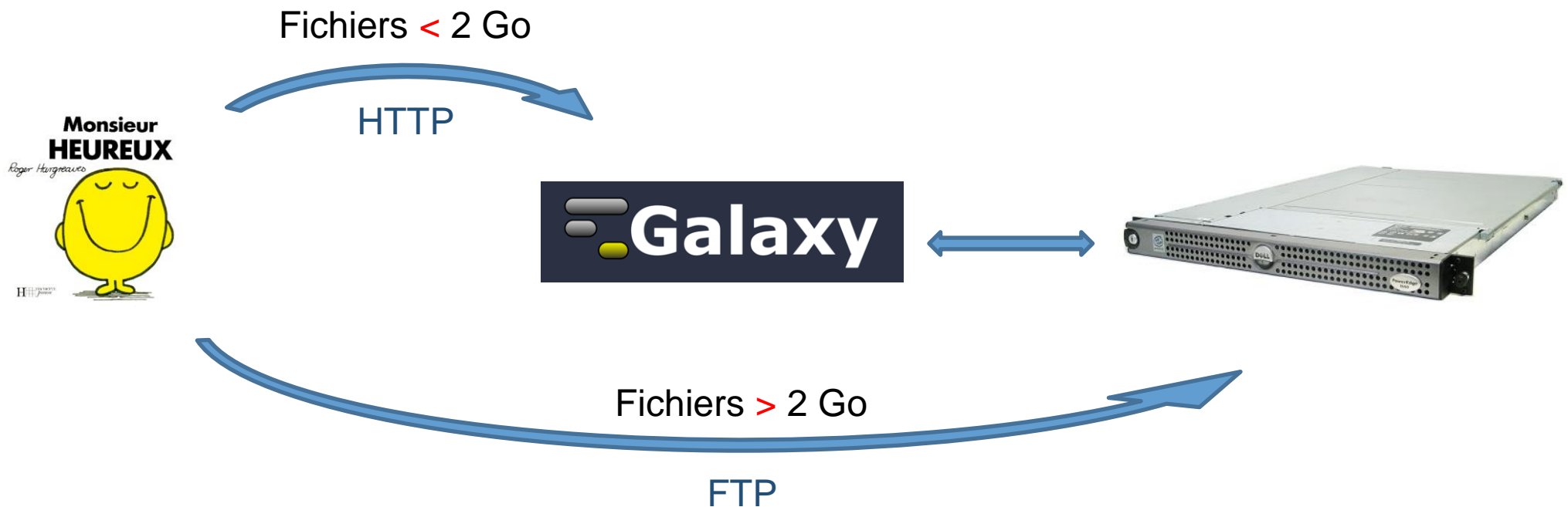
Your history is empty. Click 'Get Data' on the left pane to start

# IMPORT DES DONNÉES

## > 2 GO - PART I – CÔTÉ TRANSFERT

# Import FTP

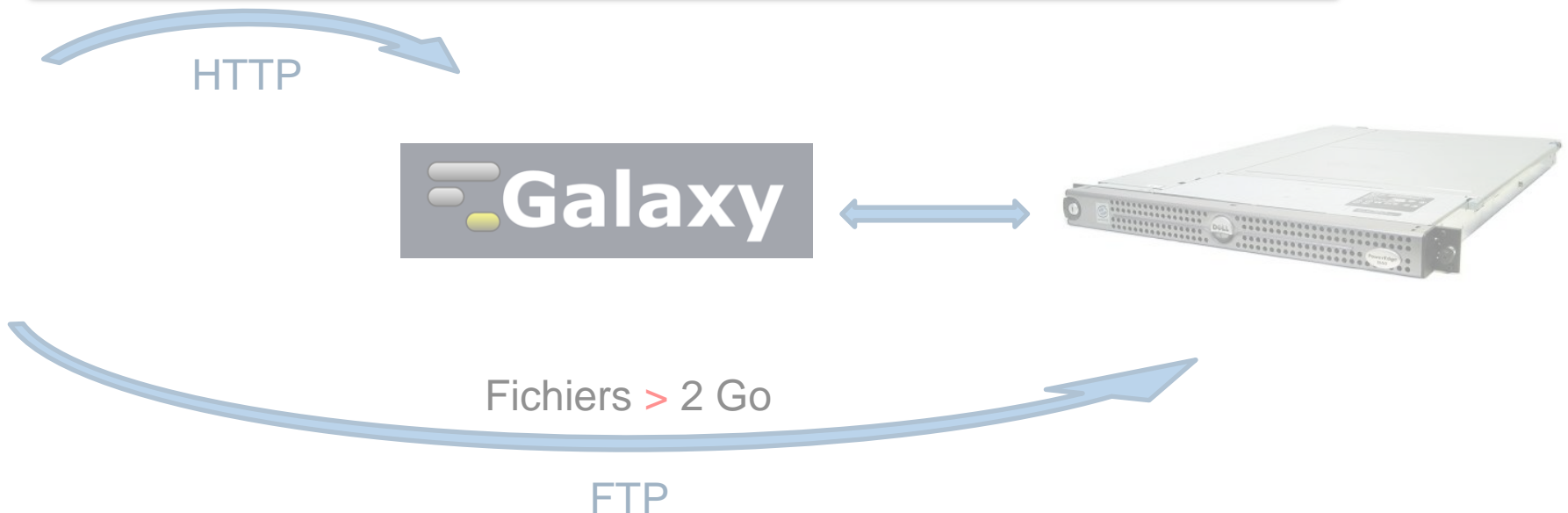
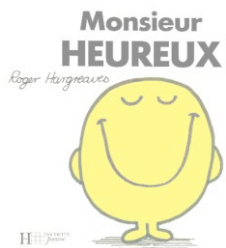
- Création d'une librairie
  - L'import des fichiers supérieurs à 2 Go doivent être déposés sur le serveur via le protocole FTP (File Transfert Protocol)



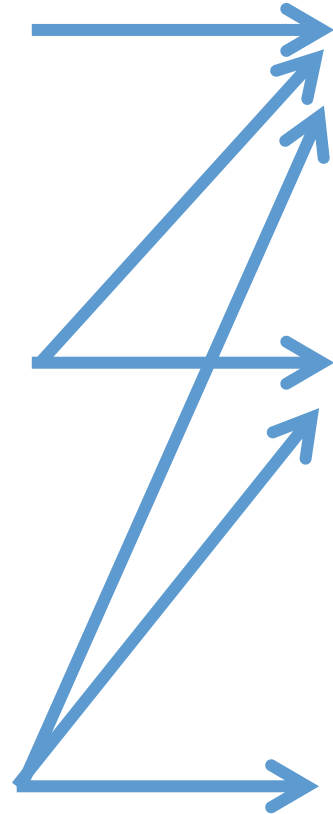
# Import FTP

- Création d'une librairie
  - L'import des fichiers supérieurs à 2 Go doivent être déposés sur le serveur via le protocole FTP (File Transfert Protocol)

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



- Les programmes de FTP

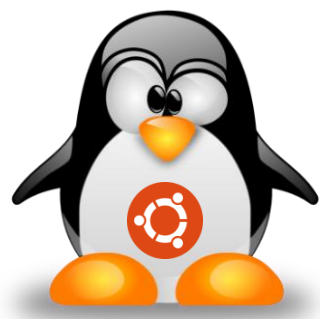


**Cyberduck**  
Libre FTP, SFTP, WebDAV & cloud storage browser for Mac & Windows.



**WinSCP**

- Les programmes de FTP



**Avoid:  
Malwares inside**



**Cyberduck**

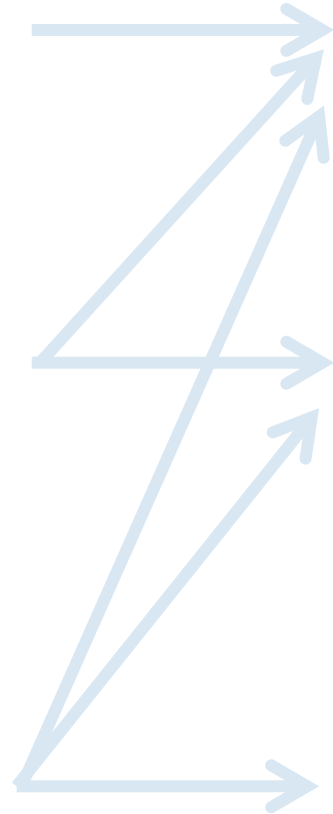
Libre FTP, SFTP, WebDAV & cloud storage browser for Mac & Windows.



**WinSCP**



- Les programmes de FTP



**FileZilla**  
The FREE FTP solution



**Cyberduck**


Libre FTP, SFTP, WebDAV & cloud storage browser for Mac & Windows.



**WinSCP**



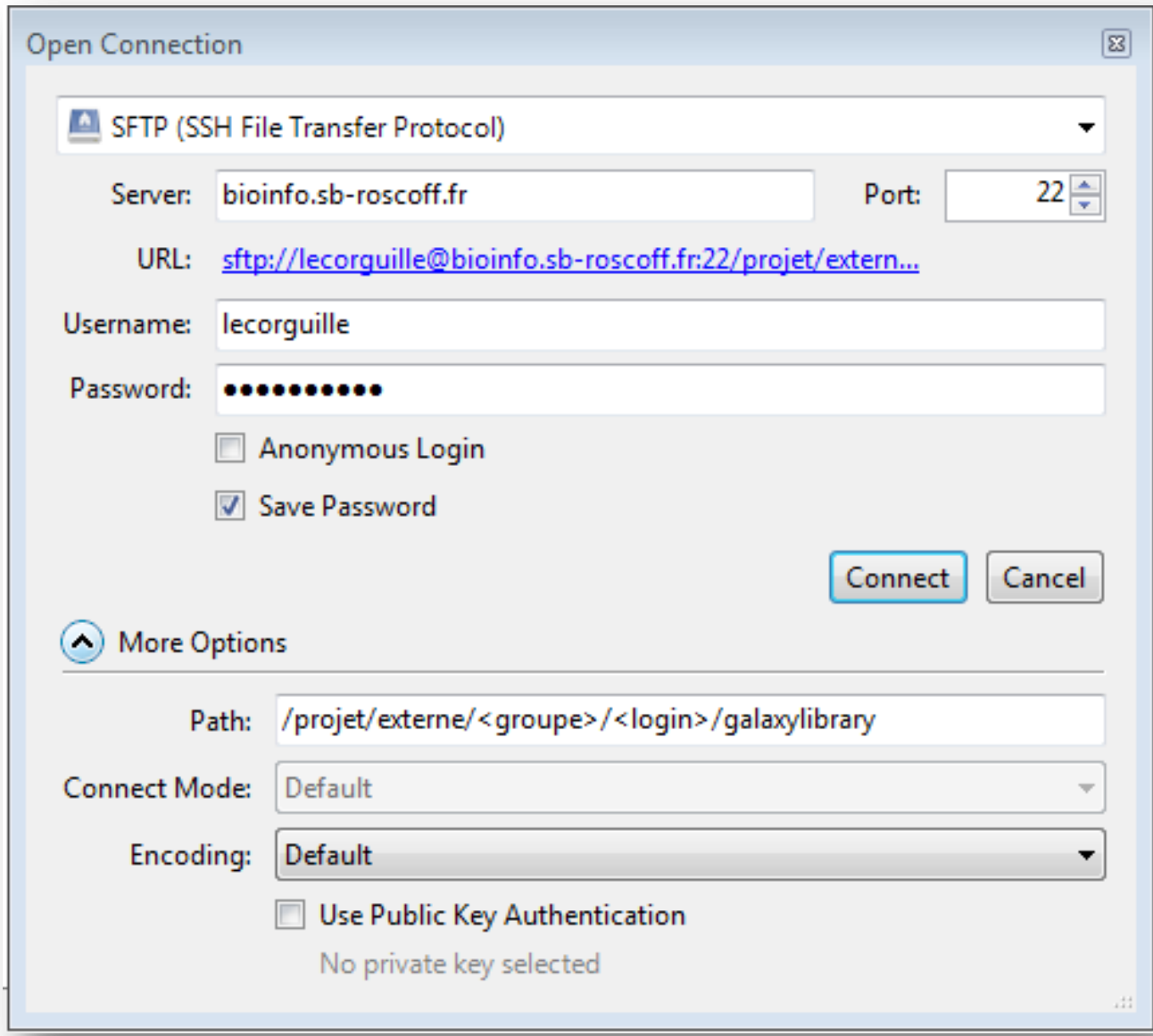
- Les paramètres

- Server / Host → bioinfo.sb-roscoff.fr
- Protocol → SFTP ou SCP
- Username / Login → 
  
- PATH →  
/projet/externe/<group>/<login>/galaxylibrary

# Import FTP

- Les

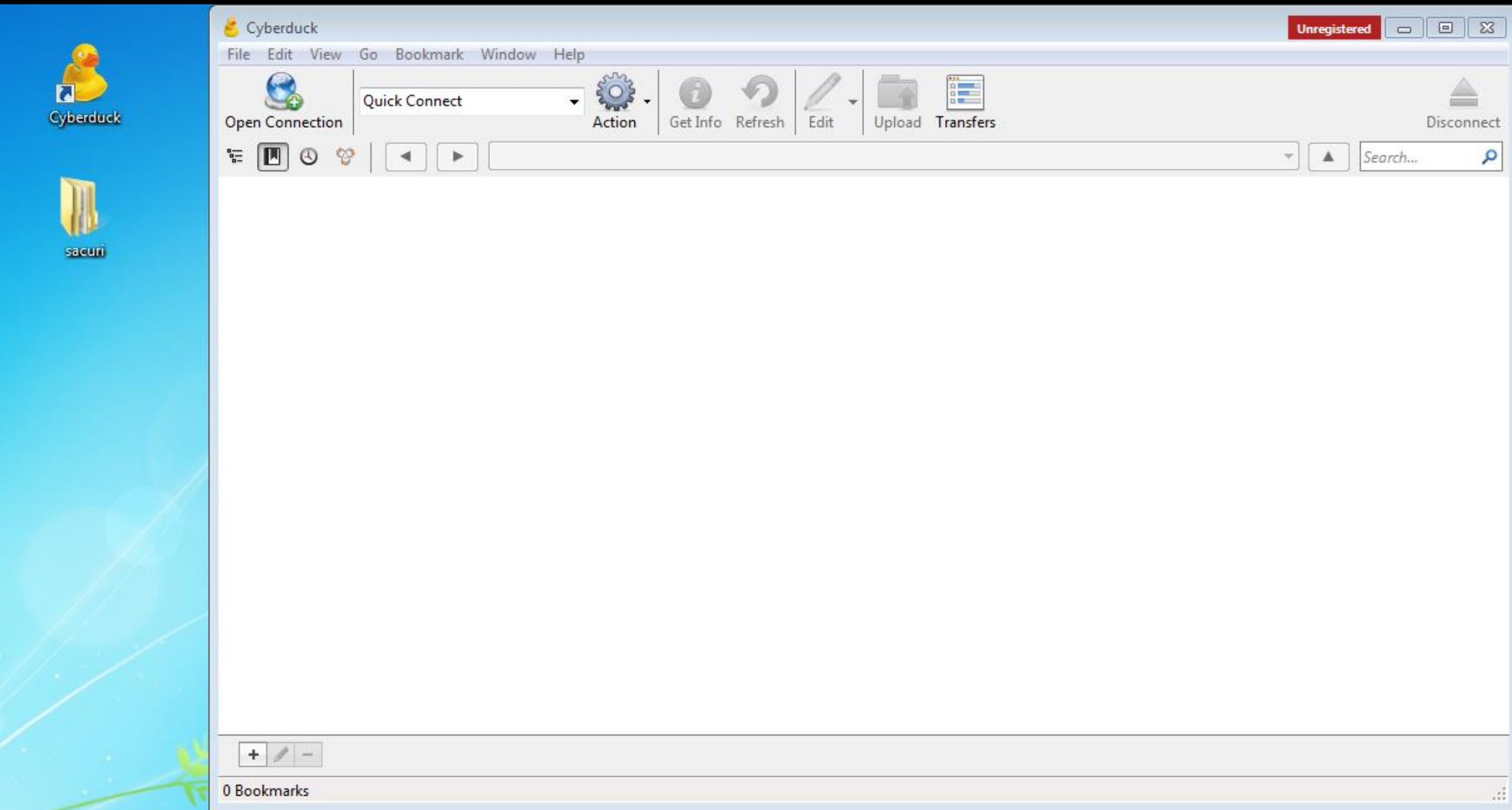
- 
- 
- 
- 



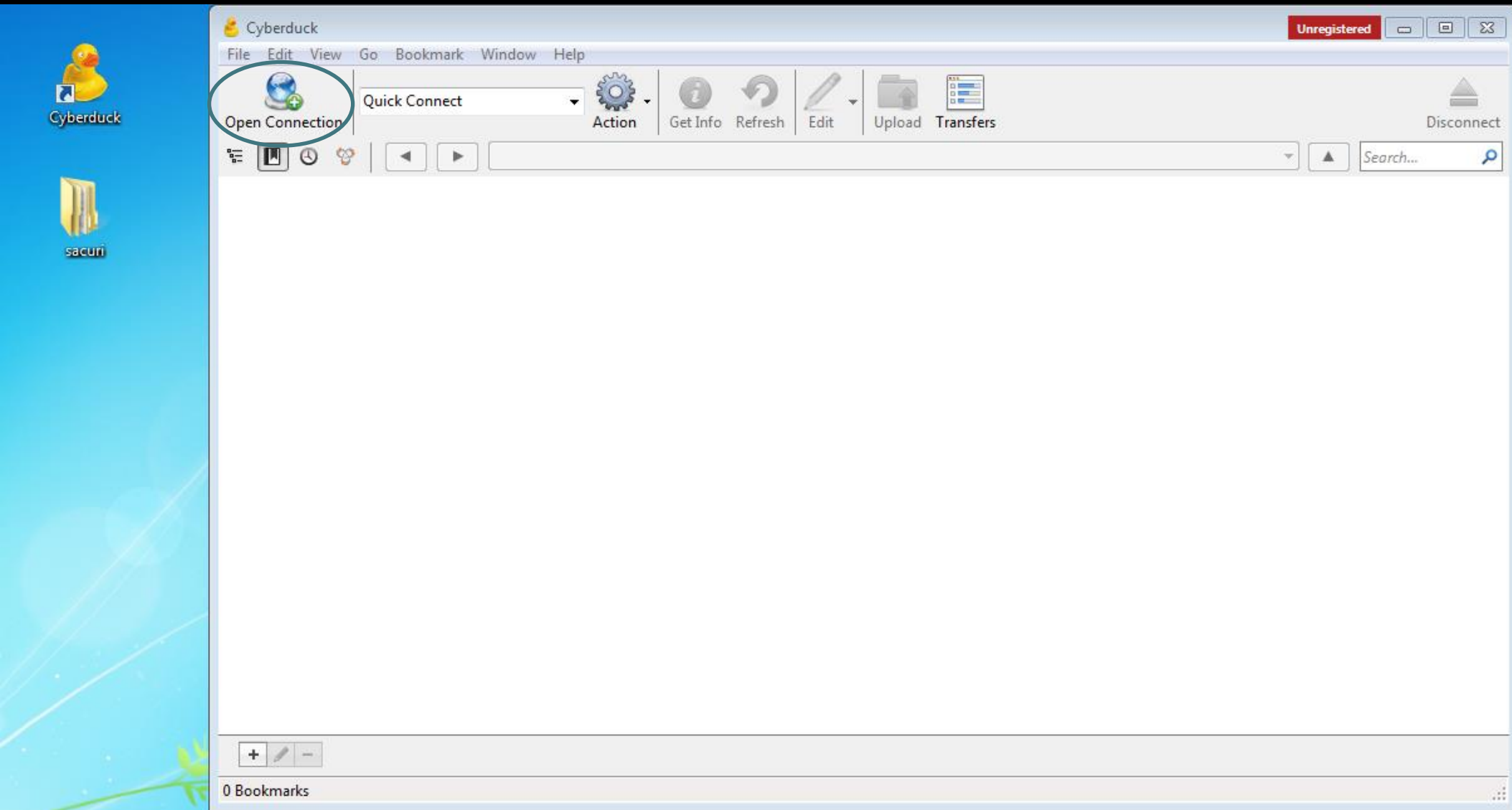
ff.fr

ary

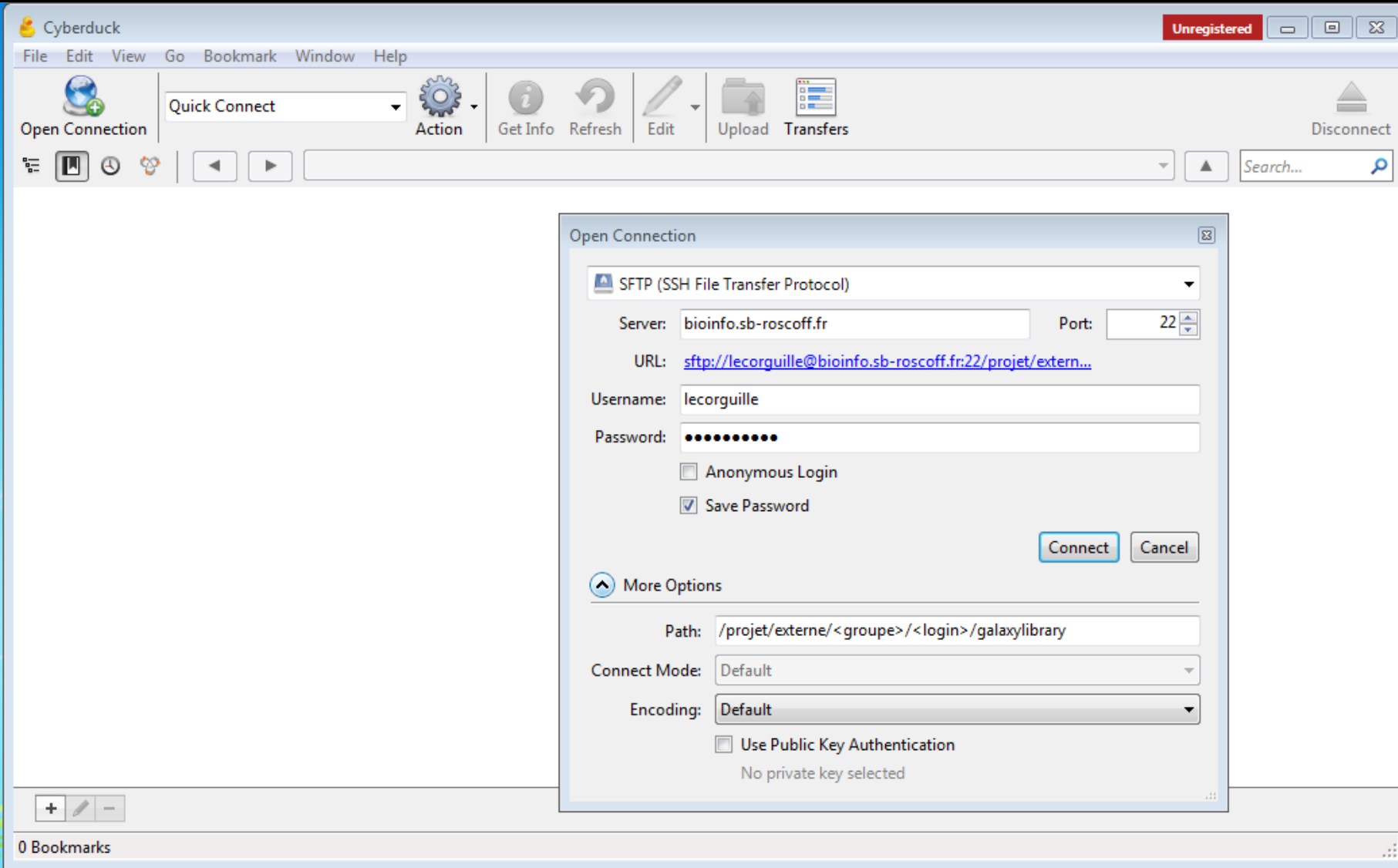
# Import d'un fichier > 2 Go



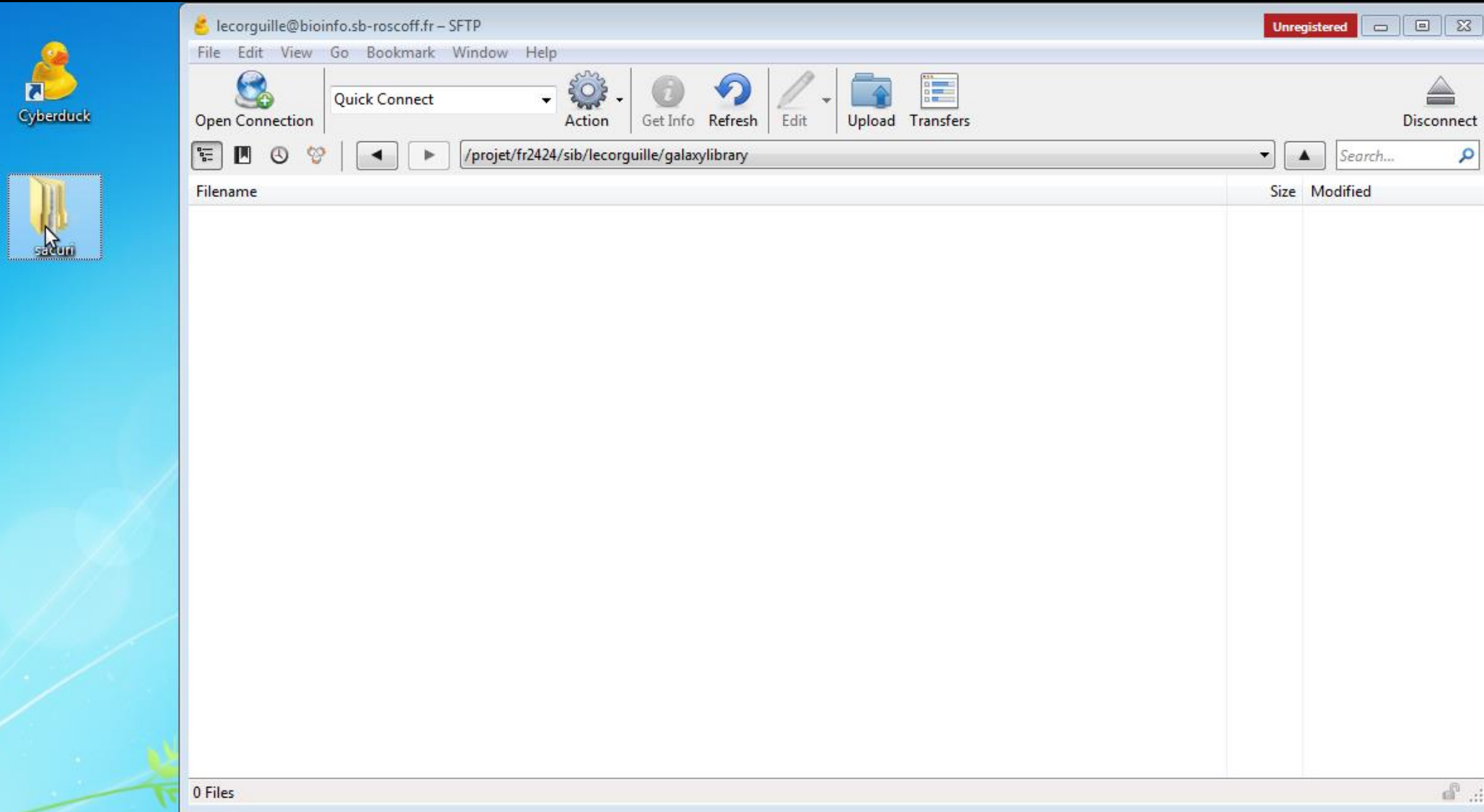
# Import d'un fichier > 2 Go



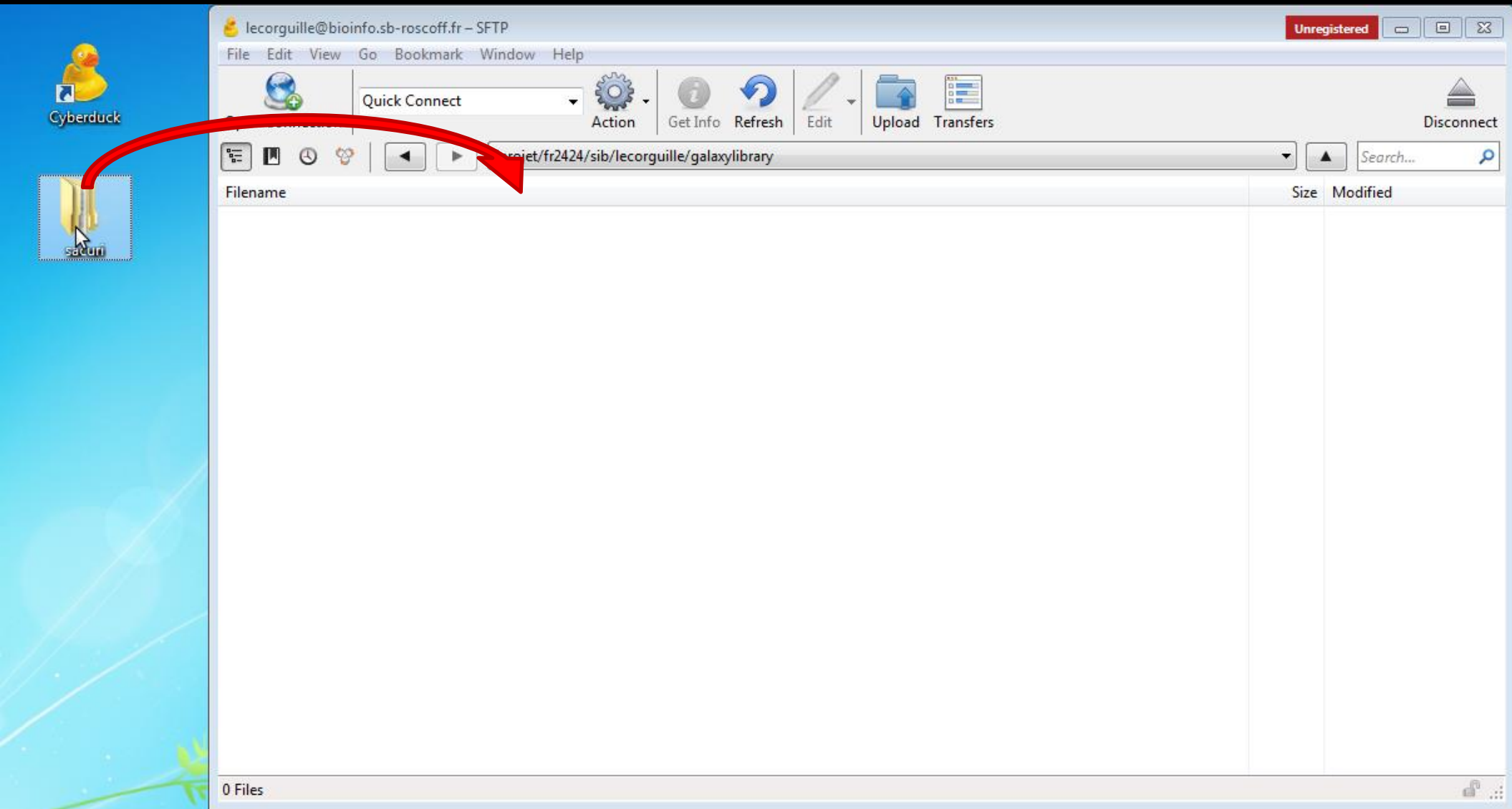
# Import d'un fichier > 2 Go



# Import d'un fichier > 2 Go



# Import d'un fichier > 2 Go



# Import d'un fichier > 2 Go

The screenshot shows the Cyberduck SFTP client interface. The main window displays the connection path `/projet/fr2424/sib/lecorguille/galaxylibrary`. A "Transfers" dialog box is open, showing the upload progress of a file named "sacuri". The progress bar indicates that 80.6 MB (84,517,228 bytes) of a 292.8 MiB file has been uploaded, at a rate of 6.8 MB/sec, with 33 seconds remaining. The file is identified as "HU\_neg\_110.mzXML". The dialog also shows the local file path `C:\Users\lecorguille\Desktop\sacuri` and the SFTP URL `sftp://bioinfo.sb-roscoff.fr...uille/galaxylibrary/sacuri`.

Filename

Size Modified

0 Files



# Import d'un fichier > 2 Go

The screenshot shows the Cyberduck SFTP client interface. The window title is "lecorguille@bioinfo.sb-roscoff.fr - SFTP". The menu bar includes File, Edit, View, Go, Bookmark, Window, and Help. The toolbar contains icons for Open Connection, Quick Connect, Action, Get Info, Refresh, Edit, Upload, Transfers, and Disconnect. The address bar shows the path "/projet/fr2424/sib/lecorguille/galaxylibrary". The main pane displays a table of files:

Filename	Size	Modified
▷ sacuri	--	5/7/2014 4:44:52 PM

The status bar at the bottom indicates "1 Files".

# IMPORT DES DONNÉES

## > 2 GO - PART II – CÔTÉ GALAXY

## Data Libraries



[Advanced Search](#)

Data library name ▾	Data library description
lecorquille	
<a href="#">RNA-seq de-novo</a>	Dataset for RNA-seq de-novo, re-ingeneered - ppericard
<a href="#">RNA-seq reference</a>	Dataset for RNA-seq with reference genome - acormier



Upload files to a data library

Browse this data library

Upload a directory of files

**Upload option:**

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

**File Format:**

**Server Directory**

Upload all files in a sub-directory of `/projetsbr/galaxy/import/user/lecorguille@sb-roscoff.fr` on the Galaxy server.

**Copy data 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 symlinks further down the line will be dereferenced.

**Convert spaces to tabs:**  
 Yes  
 Use this option if you are entering intervals by hand.

**Genome:**

**Message:**

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

**Restrict dataset access to specific roles:**

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.

Upload to library

## Data Library "Iecorguille"

Add datasets Add folder Library Actions

✓ Added 4 datasets to the library 'Iecorguille' (each is selected).

<input type="checkbox"/> Name	Message	Data type	Date uploaded	File size
<input checked="" type="checkbox"/> BlueLight.sample.paired.1.cleaned.fastq	This job is running	auto	2013-09-12	125.6 MB
<input checked="" type="checkbox"/> BlueLight.sample.paired.2.cleaned.fastq	This job is running	auto	2013-09-12	124.3 MB
<input checked="" type="checkbox"/> Dark.sample.paired.1.cleaned.fastq	This job is running	auto	2013-09-12	90.9 MB
<input checked="" type="checkbox"/> Dark.sample.paired.2.cleaned.fastq	This job is running	auto	2013-09-12	89.5 MB

For selected datasets: Import to current history Go

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

**i** 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

# ~~IMPORT DES DONNÉES~~

# Pour les ressources conséquentes et publiques

ex : swissprot au format blast  
chromosome 11 du génome humain au format bowtie2

## Faire une demande à support.abims@sb-roscoff.fr

The screenshot shows the Galaxy/ABiMS web interface for the NCBI BLAST+ blastn tool (version 0.0.17). The tool configuration is as follows:

- Nucleotide query sequence(s):** (Empty input field)
- Subject database/sequences:** BLAST Database
- Nucleotide BLAST database:** A dropdown menu is open, showing a search bar with 'nt' entered and a list of options:
  - nt
  - genbank
  - genbank Bacterial
  - genbank Environmental sampling
  - genbank EST (expressed sequence tag)
  - genbank GSS (genome survey sequence)
  - genbank HTC (high throughput cDNA sequencing)
  - genbank HTGS (high throughput genomic sequencing)
- Execute:** A blue button to run the tool.
- Note:** Database searches may take a substantial amount of time. For large input datasets it is advisable to allow overnight processing.
- What it does:** Search a *nucleotide database* using a *nucleotide query*, using the NCBI BLAST+ blastn command line tool. Algorithms include blastn, megablast, and discontinuous megablast.
- Additional Note:** You can also search against a FASTA file of subject nucleotide sequences. This is *not* advised because it is slower (only one CPU is used), but more importantly gives e-values for pairwise searches (very small e-values which will look overly significant). In most cases you should instead turn the other FASTA file into a database first using *makeblastdb* and search against that.

The right-hand side of the interface shows a 'History' panel with 'Unnamed history' and '0 bytes', along with a message: 'Your history is empty. Click 'Get Data' on the left pane to start'.

## Résumé

# IMPORT DES DONNÉES



- Résumé
  - Si ressource locale :
    - Si  $< 2$  Go :
      - Import depuis Get Data
      - Copier/Coller dans la zone de texte
    - Si  $> 2$  Go :
      - Import via le protocole FTP
  - Si ressource distante :
    - Si  $< 2$  Go :
      - Copier/Coller de l'adresse ftp://
  - Si ressource type banque publique (brute ou reformatée)
    - Demande à [support.abims@sb-roscoff.fr](mailto:support.abims@sb-roscoff.fr) pour mise à disposition

- Exercice
  - Récupérer sous Galaxy un fichier de séquence

<http://application.sb-roscoff.fr/download/fr2424/abims/lecorguille/cours/JohnDoe.fasta>

# OUTILS

## la liste des outils

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Help User Using 41%

**Tools**

search tools

**Get Data**

- Upload File from your computer

**ABiMS WORKFLOWS**

- Workflow RNA-seq de novo by ABiMS
- Workflow RNA-seq with reference by ABiMS
- Workflow 4 Metabolomics

**ABiMS TOOLS**

- Primer
- RNASeq
- InterEsil
- Statistics
- Utils
- Phylogenetics
- Debug

**COMMON TOOLS**

- Text Manipulation
- FASTA manipulation
- Join, Subtract and Group
- Filter and Sort
- NCBI BLAST+**
  - NCBI BLAST+ `blastn` Search nucleotide database with nucleotide query sequence(s)
  - NCBI BLAST+ `blastp` Search protein database with protein query sequence(s)
  - NCBI BLAST+ `blastx` Search protein database with translated nucleotide query sequence(s)
  - NCBI BLAST+ `tblastn` Search translated nucleotide database

**NCBI BLAST+ blastx (version 0.0.17)**

**Nucleotide query sequence(s):**  
1: human\_protein.fas

**Subject database/sequences:**  
FASTA file from your history (see warning note below)

**Protein FASTA file to use as database:**  
1: human\_protein.fas

**Query genetic code:**  
1. Standard

**Set expectation value cutoff:**  
0.001

**Output format:**  
Tabular (extended 24 columns)

**Advanced Options:**  
Hide Advanced Options

Execute

**Note.** Database searches may take a substantial amount of time. For large input datasets it is advisable to allow overnight processing.

**What it does**

Search a *protein database* using a *translated nucleotide query*, using the NCBI BLAST+ `blastx` command line tool.

**Note** You can also search against a FASTA file of subject protein sequences. This is *not* advised because it is slower (only one CPU is used), but more importantly gives e-values for pairwise searches (very small e-values which will look overly significant). In most cases you should instead turn the other FASTA file into a database first using `makeblastdb` and search against that.

**Output format**

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

Column	NCBI name	Description
--------	-----------	-------------

**History**

- Human protein study  
5.3 MB
- 2: chr22\_check.gff3
- 1: human\_protein.fas

60 /

## la liste des outils

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Help User Using 41%

**Tools**

blast

**NCBI BLAST+**

- NCBI BLAST+ [blastn](#) Search nucleotide database with nucleotide query sequence(s)
- NCBI BLAST+ [blastp](#) Search protein database with protein query sequence(s)
- NCBI BLAST+ [blastx](#) Search protein database with translated nucleotide query sequence(s)
- NCBI BLAST+ [tblastn](#) Search translated nucleotide database with protein query sequence(s)
- NCBI BLAST+ [tblastx](#) Search translated nucleotide database with translated nucleotide query sequence(s)
- BLAST XML to tabular Convert BLAST XML output to tabular

**Workflows**


- All workflows

**Online**

- 07-06-13: Metabolomic : Workflow 4 Metabolomics, updated to version 2.1.0 (2013\_06\_07) [i](#)
- 30-04-13: RNASeq : DESeq is now available for RNASeq expression data with reference (with gtf input).
- 26-04-13: RNASeq : DESeq is now available for denovo RNASeq expression data (without gtf input).
- 26-04-13: RNASeq : sam2counts is now available to count the reads coverage by transcrit. It's also a requirement for DESeq denovo.
- 26-04-13: Metabolomic : Workflow Metabolomic by ABiMS, updated to version 2.0.0 (2013\_04\_18) [i](#)

**Abi4**  
**AbiMS**

**Analyses and Bioinformatics for Marine Science**

 CNRS UPMC  
**Station Biologique Roscoff**

**Information**  
For any question or request for tools or account, send an email at [support.abims 'AT' sb-roscoff.fr](mailto:support.abims@AT.sb-roscoff.fr)

Galaxy is an open, web-based platform for data intensive biomedical research. The [Galaxy team](#) is a part of [BX](#) at [Penn State](#), and the [Biology and Mathematics and Computer Science](#) departments at [Emory University](#). The [Galaxy Project](#) is supported in part by [NHGRI](#), [NSF](#), [The Huck Institutes of the Life Sciences](#), [The Institute for CyberScience at Penn State](#), and [Emory University](#).

**History**

Human protein study  
5.3 MB

2: chr22\_check.gff3 [i](#) [o](#) [x](#)

1: human\_protein.fas [i](#) [o](#) [x](#)

61 /

## Exemple de formulaire :

- Sélection du fichier d'entrée – filtre sur le type de fichier (ici : fichier fasta)

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Help User Using 41%

Tools

blast

NCBI BLAST+

- NCBI BLAST+ [blastn](#) Search nucleotide database with nucleotide query sequence(s)
- NCBI BLAST+ [blastp](#) Search protein database with protein query sequence(s)
- NCBI BLAST+ [blastx](#) Search protein database with translated nucleotide query sequence(s)
- NCBI BLAST+ [tblastn](#) Search translated nucleotide database with protein query sequence(s)
- NCBI BLAST+ [tblastx](#) Search translated nucleotide database with translated nucleotide query sequence(s)
- BLAST XML to tabular Convert BLAST XML output to tabular

Workflows

- All workflows

NCBI BLAST+ blastp (version 0.0.17)

**Protein query sequence(s):**

1: human\_protein.fas

**Subject database/sequences:**

BLAST Database

**Protein BLAST database:**

nr

**Type of BLAST:**

blastp  
 blastp-short

**Set expectation value cutoff:**

0.001

**Output format:**

Tabular (extended 24 columns)

**Advanced Options:**

Hide Advanced Options

Execute

**Note.** Database searches may take a substantial amount of time. For large input datasets it is advisable to allow overnight processing.

**What it does**

Search a *protein database* using a *protein query*, using the NCBI BLAST+ blastp command line tool.

**Note.** You can also search against a FASTA file of subject protein sequences. This is *not* advised because it is slower (only one CPU is used), but more importantly gives e-values for pairwise searches (very small e-values which will look overly significant). In most cases you should instead turn the other FASTA file into a database first using *makeblastdb* and search against that.

**Output format**

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

History

Human protein study  
5.3 MB

2: chr22\_check.gff3

1: human\_protein.fas

62 /

# Exemple de formulaire :

- Aide sur l'outil

Dans la limite des stocks disponibles

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Help User Using 41%

Tools

blast

**NCBI BLAST+**

- NCBI BLAST+ [blastn](#) Search nucleotide database with nucleotide query sequence(s)
- NCBI BLAST+ [blastp](#) Search protein database with protein query sequence(s)
- NCBI BLAST+ [blastx](#) Search protein database with translated nucleotide query sequence(s)
- NCBI BLAST+ [tblastn](#) Search translated nucleotide database with protein query sequence(s)
- NCBI BLAST+ [tblastx](#) Search translated nucleotide database with translated nucleotide query sequence(s)
- BLAST XML to tabular Convert BLAST XML output to tabular

Workflows

- All workflows

**NCBI BLAST+ blastp (version 0.0.17)**

**Protein query sequence(s):**  
1: human\_protein.fas

**Subject database/sequences:**  
BLAST Database

**Protein BLAST database:**  
nr

**Type of BLAST:**  
 blastp  
 blastp-short

**Set expectation value cutoff:**  
0.001

**Output format:**  
Tabular (extended 24 columns)

**Advanced Options:**  
Hide Advanced Options

**Execute**

**Note.** Database searches may take a substantial amount of time. For large input datasets it is advisable to allow overnight processing.

**What it does**  
Search a *protein database* using a *protein query*, using the NCBI BLAST+ blastp command line tool.

**Note.** You can also search against a FASTA file of subject protein sequences. This is *not* advised because it is slower (only one CPU is used), but more importantly gives e-values for pairwise searches (very small e-values which will look overly significant). In most cases you should instead turn the other FASTA file into a database first using *makeblastdb* and search against that.

**Output format**  
Because Galaxy focuses on processing tabular data, the default output of this tool is tabular. The standard BLAST+ tabular output contains 12 columns:

History

- Human protein study  
5.3 MB
- 2: chr22\_check.gff3
- 1: human\_protein.fas

63 /

# Exemple de formulaire :

• Ici : possibilité de choisir une banque personnelle dans l'historique ... mais j'en ai pas !

The screenshot displays the Galaxy/ABiMS interface for the NCBI BLAST+ blastp tool. The main configuration area includes fields for 'Protein query sequence(s)' (set to '1: human\_protein.fas'), 'Subject database/sequences' (highlighted with a red box and set to 'BLAST database from your history'), 'Protein BLAST database', 'Type of BLAST' (radio buttons for 'blastp' and 'blastp-short'), 'Set expectation value cutoff' (input field with '0.001'), 'Output format' (dropdown set to 'Tabular (extended 24 columns)'), and 'Advanced Options' (button for 'Hide Advanced Options'). An 'Execute' button is located at the bottom of the configuration area. A warning note states: 'Database searches may take a substantial amount of time. For large input datasets it is advisable to allow overnight processing.' Below this, a section titled 'What it does' explains the tool's function: 'Search a protein database using a protein query, using the NCBI BLAST+ blastp command line tool.' A second warning note states: 'You can also search against a FASTA file of subject protein sequences. This is not advised because it is slower (only one CPU is used), but more importantly gives e-values for pairwise searches (very small e-values which will look overly significant). In most cases you should instead turn the other FASTA file into a database first using makeblastdb and search against that.' The 'Output format' section notes: 'Because Galaxy focuses on processing tabular data, the default output of this tool is tabular. The standard BLAST+ tabular output contains 12 columns:'. The right sidebar shows a 'History' panel with a list of jobs: 'Human protein study' (5.3 MB), '2: chr22\_check.gff3', and '1: human\_protein.fas' (selected).



## Exemple de formulaire :

- Possibilité de choisir le format du fichier de sortie

Dans la limite des stocks disponibles

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Help User Using 41%

Tools

blast

**NCBI BLAST+**

- NCBI BLAST+ [blastn](#) Search nucleotide database with nucleotide query sequence(s)
- NCBI BLAST+ [blastp](#) Search protein database with protein query sequence(s)
- NCBI BLAST+ [blastx](#) Search protein database with translated nucleotide query sequence(s)
- NCBI BLAST+ [tblastn](#) Search translated nucleotide database with protein query sequence(s)
- NCBI BLAST+ [tblastx](#) Search translated nucleotide database with translated nucleotide query sequence(s)
- BLAST XML to tabular Convert BLAST XML output to tabular

Workflows

- All workflows

**NCBI BLAST+ blastp (version 0.0.17)**

**Protein query sequence(s):**  
1: human\_protein.fas

**Subject database/sequences:**  
BLAST Database

**Protein BLAST database:**  
nr

**Type of BLAST:**  
 blastp  
 blastp-short

**Set expectation value cutoff:**  
0.001

**Output format:**  
Tabular (extended 24 columns)

**Advanced Options:**  
Hide Advanced Options

**Execute**

**Note.** Database searches may take a substantial amount of time. For large input datasets it is advisable to allow overnight processing.

**What it does**  
Search a *protein database* using a *protein query*, using the NCBI BLAST+ blastp command line tool.

**Note** You can also search against a FASTA file of subject protein sequences. This is *not* advised because it is slower (only one CPU is used), but more importantly gives e-values for pairwise searches (very small e-values which will look overly significant). In most cases you should instead turn the other FASTA file into a database first using *makeblastdb* and search against that.

**Output format**  
Because Galaxy focuses on processing tabular data, the default output of this tool is tabular. The standard BLAST+ tabular output contains 12 columns:

History

Human protein study  
5.3 MB

2: chr22\_check.gff3

1: human\_protein.fas

# Exemple de formulaire :

• Possibilité d'accéder à des options avancées = accessibilité / ergonomie

Dans la limite des stocks disponibles

The screenshot shows the Galaxy/ABiMS interface for the NCBI BLAST+ blastp tool (version 0.0.17). The interface includes a top navigation bar with 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User' menus. A 'Tools' sidebar on the left lists various BLAST-related tools. The main configuration area contains the following fields:

- Protein query sequence(s):** 1: human\_protein.fas
- Subject database/sequences:** BLAST Database
- Protein BLAST database:** nr
- Type of BLAST:** blastp (selected), blastp-short
- Set expectation value cutoff:** 0.001
- Output format:** Tabular (extended 24 columns)
- Advanced Options:** A red-bordered box containing:
  - Show Advanced Options (dropdown)
  - Filter out low complexity regions (with SEG):
  - Scoring matrix: BLOSUM62 (default)
  - Maximum hits to show: 0 (with note: Use zero for default limits)
  - Word size for wordfinder algorithm: 0 (with note: Use zero for default, otherwise minimum 2.)
  - Should the query and subject define(s) be parsed?:  (with note: This affects the formatting of the query/subject ID strings)

On the right, a 'History' panel shows a list of jobs: 'Human protein study' (5.3 MB), '2: chr22\_check.gff3', and '1: human\_protein.fas'. An 'Execute' button is located at the bottom of the tool configuration area.

- Les status

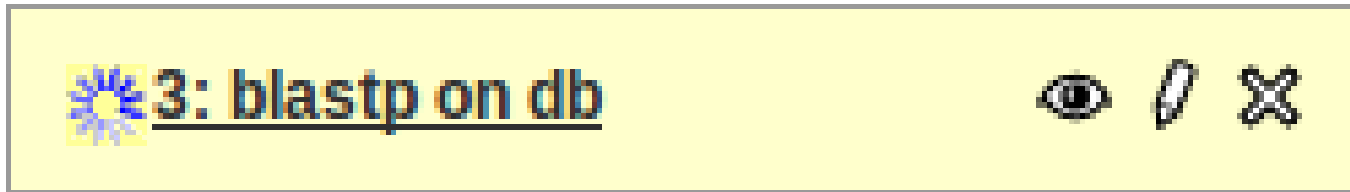


Job en attente de soumission

= le job est dans la « queue » de l'ordonnanceur

La durée de ce statut dépend du nombre de jobs actuellement en attente et du nombre de cpu demandé

- Les status



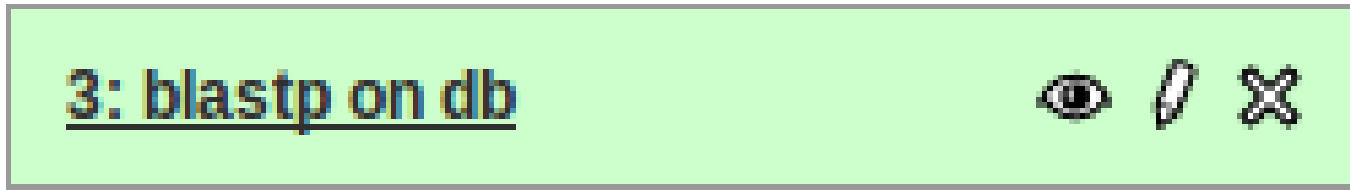
## Job en cours d'exécution

= le job tourne actuellement sur le cluster de calcul

La durée de ce status dépend complètement de la nature du job et de la puissance de calcul allouée

Certains programmes vont pouvoir tourner sur plusieurs processeurs et disposer de 4, 8 ou 16 Go de RAM. Et d'autres sont mono-CPU.

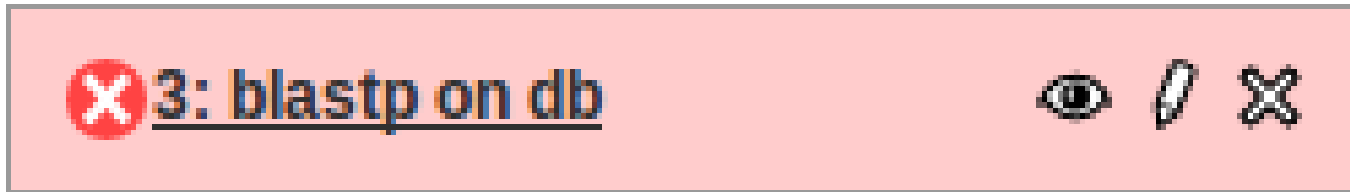
- Les status



Job terminé

Son statut est OK mais des warnings ou des erreurs peuvent se cacher derrière. Ah hum !

- Les status



Job terminé mais en erreur

= le programme renvoie une erreur

Un programme si il est bien développé renvoie un code d'erreur :

- = 0 si tout s'est bien terminé
- > 0 si il a rencontré une erreur

Le programme explique souvent d'où vient l'erreur

et parfois ... pas.

- Les status



Job terminé mais en erreur

Les sources d'erreur :

- L'utilisateur :P
- Mauvaise utilisation : fichier d'entrée ou format ou option
- Mauvais portage du programme sous Galaxy ... désolé :/
- Plantage non anticipé du programme

## Exemple d'erreur

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 19.1 GB

Tools search tools

**Get Data**

ABIMS WORKFLOWS

[Workflow RNA-seq de novo by ABiMS](#)

[Workflow RNA-seq with reference by ABiMS](#)

[Workflow 4 Metabolomics](#)

ABIMS TOOLS

[Primer](#)

[RNASeq](#)

[InterEsil](#)

[Statistics](#)

[Utils](#)

[Phylogenetics](#)

[Debug](#)

COMMON TOOLS

[Text Manipulation](#)

[FASTA manipulation](#)

[Join, Subtract and Group](#)

[Filter and Sort](#)

[Graphics](#)

[NCBI BLAST+](#)

[NGS: QC and manipulation](#)

[NGS: RNA Analysis](#)

[NGS: Mapping](#)

[NGS: Picard \(beta\)](#)

[NGS: SAM Tools](#)

[NGS: GATK Tools \(beta\)](#)

[SVDetect](#)

[VarScan](#)

[Searching sequence tools](#)


**Online**

**i**

- 07-06-13: Metabolomic : Workflow 4 Metabolomics, updated to version 2.1.0 (2013\_06\_07) [↓](#)
- 30-04-13: RNASeq : DESeq is now available for RNASeq expression data with reference (with gtf input).
- 26-04-13: RNASeq : DESeq is now available for denovo RNASeq expression data (without gtf input).
- 26-04-13: RNASeq : sam2counts is now available to count the reads coverage by transcrit. It's also a requirement for DESeq denovo.
- 26-04-13: Metabolomic : Workflow Metabolomic by ABiMS, updated to version 2.0.0 (2013\_04\_18) [↓](#)

**AbiMS**

Analyses and Bioinformatics for Marine Science

 CNRS UPMC  
Station Biologique  
Roscoff

**i** **Information**  
For any question or request for tools or account, send an email at [support.abims@sb-roscoff.fr](mailto:support.abims@sb-roscoff.fr)

Galaxy is an open, web-based platform for data intensive biomedical research. The [Galaxy team](#) is a part of [BX](#) at [Penn State](#), and the [Biology and Mathematics and Computer Science](#) departments at [Emory University](#). The [Galaxy Project](#) is supported in part by [NHGRI](#), [NSF](#), [The Huck Institutes of the Life Sciences](#), [The Institute for CyberScience at Penn State](#), and [Emory University](#).

**History**

**Copper Stress v3**  
133.9 MB

**XZ:** [mzXML\\_copper\\_stress.group.retcor.group.fillPeaks.annotateDiffreport.data\\_matrix.tsv\\_anova\\_pvalue.tabular](#)

**6:** [mzXML\\_copper\\_stress.group.retcor.group.fillPeaks.annotateDiffreport.Rdata](#)

**5:** [mzXML\\_copper\\_stress.group.retcor.group.fillPeaks.RData](#)

**4:** [mzXML\\_copper\\_stress.group.retcor.group.RData](#)

**3:** [mzXML\\_copper\\_stress.RData](#)

**2:** [sampleInfo.tab](#)

**1:** [mzXML\\_copper\\_stress.ms.zip](#)

72 /







# Exemple d'erreur

Galaxy / ABiMS

Analyze Data | Workflow | Shared Data | Visualization | Admin | Help | User

Using 18.1 GB

### Tools

Get Data

ABiMS WORKFLOWS

- [Workflow RNA-seq de novo by ABiMS](#)
- [Workflow RNA-seq with reference by ABiMS](#)
- [Workflow 4 Metabolomics](#)

ABiMS TOOLS

- [Primer](#)
- [RNASeq](#)
- [InterEsil](#)
- [Statistics](#)
- [Utils](#)
- [Phylogenetics](#)
- [Debug](#)

COMMON TOOLS

- [Text Manipulation](#)
- [FASTA manipulation](#)
- [Join, Subtract and Group](#)
- [Filter and Sort](#)
- [Graphics](#)
- [NCBI BLAST+](#)
- [NGS: QC and manipulation](#)
- [NGS: RNA Analysis](#)
- [NGS: Mapping](#)
- [NGS: Picard \(beta\)](#)
- [NGS: SAM Tools](#)
- [NGS: GATK Tools \(beta\)](#)
- [SVDetect](#)
- [VarScan](#)
- [Muscle](#)
- [RAxML](#)

NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA

**Report this error to the Galaxy Team**

The Galaxy team regularly reviews errors that occur in the application. However, if you would like to provide additional information (such as what you were trying to do when the error occurred) and a contact e-mail address, we will be better able to investigate your problem and get back to you.

**Error Report**

Your email

**Message**

Report

### History

**Copper Stress v3**  
133.9 MB

**XZ:** mzXML copper\_stress.group.retcor.roup.fillPeaks.annotateDiffreport.data\_matrix.tsv anova pvalue.tabular error  
An error occurred with this dataset:  
*Fatal error: Exit code 10 () ERROR: There is a problem with the group of condition (presence of NA). You may need to use change the mode (column/row) Current groups : NA*

**6:** mzXML copper\_stress.group.retcor.roup.fillPeaks.annotateDiffreport.Rdata a

**5:** mzXML copper\_stress.group.retcor.roup.fillPeaks.RData

**4:** mzXML copper\_stress.group.retcor.roup.RData

**3:** mzXML copper\_stress.RData

**2: sampleInfo.tab**

**1:** mzXML copper\_stress.ms.zip data  
format: ms\_zip, database: ?  
uploaded ms\_zip file 75 /

# HISTORIQUE

Contient aussi bien les entrées que les sorties

Galaxy / ABiMS
Analyze Data Workflow Shared Data Visualization Admin Help User
Using 18.0 GB

Tools

**Get Data**

ABiMS WORKFLOWS

[Workflow RNA-seq de novo by ABiMS](#)

[Workflow RNA-seq with reference by ABiMS](#)

[Workflow 4 Metabolomics](#)

ABiMS TOOLS

[Primer](#)

[RNASeq](#)

[InterEsil](#)

[Statistics](#)

[Utils](#)

[Phylogenetics](#)

[Debug](#)

COMMON TOOLS

[Text Manipulation](#)

[FASTA manipulation](#)

[Join, Subtract and Group](#)

[Filter and Sort](#)

[Graphics](#)

[NCBI BLAST+](#)

[NGS: QC and manipulation](#)

[NGS: RNA Analysis](#)

[NGS: Mapping](#)

[NGS: Picard \(beta\)](#)

[NGS: SAM Tools](#)

[NGS: GATK Tools \(beta\)](#)

[SVDetect](#)

[VarScan](#)


[Muscle](#)

Searchina sequence tools


✓ Online

i

- 07-06-13: Metabolomic : Workflow 4 Metabolomics, updated to version 2.1.0 (2013\_06\_07) i
- 30-04-13: RNASeq : DESeq is now available for RNASeq expression data with reference (with gtf input).
- 26-04-13: RNASeq : DESeq is now available for denovo RNASeq expression data (without gtf input).
- 26-04-13: RNASeq : sam2counts is now available to count the reads coverage by transcrit. It's also a requirement for DESeq denovo.
- 26-04-13: Metabolomic : Workflow Metabolomic by ABiMS, updated to version 2.0.0 (2013\_04\_18) i



## Analyses and Bioinformatics for Marine Science



CNRS UPMC

### Station Biologique Roscoff

i **Information**  
 For any question or request for tools or account, send an email at [support.abims 'AT' sb-roscoff.fr](mailto:support.abims@at.sb-roscoff.fr)

Galaxy is an open, web-based platform for data intensive biomedical research. The [Galaxy team](#) is a part of [BX](#) at [Penn State](#), and the [Biology and Mathematics and Computer Science](#) departments at [Emory University](#). The [Galaxy Project](#) is supported in part by [NHGRI](#), [NSF](#), [The Huck Institutes of the Life Sciences](#), [The Institute for CyberScience at Penn State](#), and [Emory University](#).

**History**



**Human chr22 proteome study**  
19.1 MB



**3: blastp on db**




**2: chr22\_check.gff3**




**1: human\_protein.fas**




## Renommer, tagger et annoter

History  



Unnamed history  
19.1 MB  



3: blastp on db   




2: chr22\_check.gff3   




1: human\_protein.fas   






History  

Human chr22 proteome study  
19.1 MB  

3: blastp on db   

2: chr22\_check.gff3   

1: human\_protein.fas   



History  

Human chr22 proteome study  
19.1 MB  

Tags:  
     



Annotation:  
 Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

3: blastp on db   

2: chr22\_check.gff3   

1: human\_protein.fas   

## Information sur l'objet :

- Code erreur
- Sorties standard / error

Galaxy / ABiMS

[Analyze Data](#)
[Workflow](#)
[Shared Data](#)
[Visualization](#)
[Admin](#)
[Help](#)
[User](#)
Using 18.0 GB

### Tool: NCBI BLAST+ blastp

Name:	blastp on db
Created:	Aug 09, 2013
Filesize:	13.9 MB
Dbkey:	?
Format:	tabular
Galaxy Tool Version:	0.0.17
Tool Version:	blastp: 2.2.27+ Package: blast 2.2.27, build Sep 11 2012 10:14:26
Tool Standard Output:	<a href="#">stdout</a>
Tool Standard Error:	<a href="#">stderr</a>
Tool Exit Code:	0
API ID:	8a4a8f9f3df4a393
Full Path:	/w/galaxy/dev/galaxy-dist/database/files/001/dataset_1534.dat

Input Parameter	Value	Note for rerun
Protein query sequence(s)	1: human_protein.fas	
Subject database/sequences	db	
Protein BLAST database	uniprot_swissprot	
histdb		
subject		
Type of BLAST	blastp	
Set expectation value cutoff	0.0001	
Output format	Tabular (extended 24 columns)	
Advanced Options	basic	

### Inheritance Chain

**blastp on db**

↑

'blastp on db' in Blastp

### History

**Human chr22 proteome study**  
 19.1 MB

**3: blastp on db**  
 21,556 lines  
 format: tabular, database: ?

1	2	3	4	5	6
sp P31946 1433B_HUMAN	sp A4K2U9 1433B_PONAB	100.00	246		
RYLSEVASGDNKQTTVSNSQQAYQEAFEISKKEMQPTHPIRLGLALNFSVFYYEILNSF					
DNKQTTVSNSQQAYQEAFEISKKEMQPTHPIRLGLALNFSVFYYEILNSPEKACSLAKT					
sp P31946 1433B_HUMAN	sp Q4R572 1433B_MACFA	100.00	246		
RYLSEVASGDNKQTTVSNSQQAYQEAFEISKKEMQPTHPIRLGLALNFSVFYYEILNSF					
DNKQTTVSNSQQAYQEAFEISKKEMQPTHPIRLGLALNFSVFYYEILNSPEKACSLAKT					

**2: chr22\_check.gff3**

**1: human\_protein.fas**

79 /



## Visualisation de l'objet

- Tableau
- Image
- HTML

Galaxy / ABiMS
Analyze Data Workflow Shared Data Visualization Admin Help User
Using 18.0 GB

sp P31946 1433B_HUMAN	sp A4K2U9 1433B_PONAB	100.00	246	0	0	1	246	1	246	0.0	508	sp A4K2U9 1433B_PO
sp P31946 1433B_HUMAN	sp Q4R572 1433B_MACFA	100.00	246	0	0	1	246	1	246	0.0	508	sp Q4R572 1433B_MA
sp P31946 1433B_HUMAN	sp P31946 1433B_HUMAN	100.00	246	0	0	1	246	1	246	0.0	508	sp P31946 1433B_HU
sp P31946 1433B_HUMAN	sp P68250 1433B_BOVIN	99.59	246	1	0	1	246	1	246	0.0	506	sp P68250 1433B_BO
sp P31946 1433B_HUMAN	sp Q9CQV8 1433B_MOUSE	98.78	246	3	0	1	246	1	246	3e-179	499	sp Q9CQV8 1433B_M
sp P31946 1433B_HUMAN	sp P35213 1433B_RAT	98.37	246	4	0	1	246	1	246	2e-178	497	sp P35213 1433B_RA
sp P31946 1433B_HUMAN	sp Q5ZLQ6 1433B_CHICK	97.95	244	5	0	3	246	1	244	3e-177	494	sp Q5ZLQ6 1433B_CH
sp P31946 1433B_HUMAN	sp Q6UFZ9 143B1_ONCMY	91.80	244	20	0	3	246	1	244	1e-164	462	sp Q6UFZ9 143B1_ON
sp P31946 1433B_HUMAN	sp Q5PRD0 143BA_DANRE	91.39	244	21	0	3	246	1	244	3e-164	461	sp Q5PRD0 143BA_D
sp P31946 1433B_HUMAN	sp Q5XHK2 143BA_XENLA	88.11	244	29	0	3	246	1	244	2e-159	449	sp Q5XHK2 143BA_XE
sp P31946 1433B_HUMAN	sp Q8AVQ3 143BB_XENLA	87.70	244	30	0	3	246	1	244	6e-159	447	sp Q8AVQ3 143BB_XE
sp P31946 1433B_HUMAN	sp Q5XGC8 1433B_XENTR	88.52	244	28	0	3	246	1	244	6e-159	447	sp Q5XGC8 1433B_XE
sp P31946 1433B_HUMAN	sp P63102 1433Z_RAT	88.02	242	29	0	3	244	1	242	2e-157	444	sp P63102 1433Z_RA
sp P31946 1433B_HUMAN	sp P63101 1433Z_MOUSE	88.02	242	29	0	3	244	1	242	2e-157	444	sp P63101 1433Z_MO
sp P31946 1433B_HUMAN	sp P29361 1433Z_SHEEP	87.60	242	30	0	3	244	1	242	1e-156	442	sp P29361 1433Z_SHE
sp P31946 1433B_HUMAN	sp Q5R651 1433Z_PONAB	87.60	242	30	0	3	244	1	242	3e-156	441	sp Q5R651 1433Z_PO
sp P31946 1433B_HUMAN	sp P63104 1433Z_HUMAN	87.60	242	30	0	3	244	1	242	3e-156	441	sp P63104 1433Z_HU
sp P31946 1433B_HUMAN	sp P63103 1433Z_BOVIN	87.60	242	30	0	3	244	1	242	3e-156	441	sp P63103 1433Z_BO
sp P31946 1433B_HUMAN	sp Q5ZKC9 1433Z_CHICK	87.19	242	31	0	3	244	1	242	7e-156	440	sp Q5ZKC9 1433Z_CH
sp P31946 1433B_HUMAN	sp Q7T356 143BB_DANRE	88.52	244	26	1	3	246	1	242	2e-155	439	sp Q7T356 143BB_DA
sp P31946 1433B_HUMAN	sp P29309 1433_XENLA	87.23	235	30	0	12	246	1	235	6e-152	429	sp P29309 1433_XENI
sp P31946 1433B_HUMAN	sp Q6UFZ8 143B2_ONCMY	88.41	233	27	0	3	235	1	233	3e-150	426	sp Q6UFZ8 143B2_ON
sp P31946 1433B_HUMAN	sp Q6P4Z5 1433Z_XENTR	84.71	242	37	0	3	244	1	242	4e-150	426	sp Q6P4Z5 1433Z_XE
sp P31946 1433B_HUMAN	sp Q91896 1433Z_XENLA	83.47	242	40	0	3	244	1	242	3e-148	421	sp Q91896 1433Z_XE
sp P31946 1433B_HUMAN	sp Q5ZMD1 1433T_CHICK	82.23	242	43	0	3	244	1	242	1e-147	419	sp Q5ZMD1 1433T_CH
sp P31946 1433B_HUMAN	sp Q5RFJ2 1433T_PONAB	81.82	242	44	0	3	244	1	242	2e-146	416	sp Q5RFJ2 1433T_PO
sp P31946 1433B_HUMAN	sp P27348 1433T_HUMAN	81.82	242	44	0	3	244	1	242	2e-146	416	sp P27348 1433T_HU
sp P31946 1433B_HUMAN	sp Q3SZI4 1433T_BOVIN	81.82	242	44	0	3	244	1	242	2e-146	416	sp Q3SZI4 1433T_BO
sp P31946 1433B_HUMAN	sp Q52M98 1433T_XENLA	81.82	242	44	0	3	244	1	242	2e-146	416	sp Q52M98 1433T_XE
sp P31946 1433B_HUMAN	sp P68255 1433T_RAT	81.82	242	44	0	3	244	1	242	2e-146	416	sp P68255 1433T_RA
sp P31946 1433B_HUMAN	sp Q6Q6X0 1433T_RABIT	81.82	242	44	0	3	244	1	242	2e-146	416	sp Q6Q6X0 1433T_RA
sp P31946 1433B_HUMAN	sp P68254 1433T_MOUSE	81.82	242	44	0	3	244	1	242	2e-146	416	sp P68254 1433T_MO
sp P31946 1433B_HUMAN	sp Q2F637 1433Z_BOMMO	79.92	244	49	0	1	244	1	244	3e-143	408	sp Q2F637 1433Z_BO
sp P31946 1433B_HUMAN	sp P29310 1433Z_DROME	79.42	243	50	0	2	244	3	245	3e-142	405	sp P29310 1433Z_DR
sp P31946 1433B_HUMAN	sp Q1HR36 1433Z_AEDAE	78.60	243	52	0	2	244	3	245	2e-140	401	sp Q1HR36 1433Z_AE
sp P31946 1433B_HUMAN	sp Q20655 14332_CAEEL	78.78	245	52	0	1	245	1	245	7e-140	399	sp Q20655 14332_CAI
sp P31946 1433B_HUMAN	sp P41932 14331_CAEEL	78.10	242	48	2	7	244	7	247	1e-131	378	sp P41932 14331_CAI

**History**

**Human chr22 proteome study**  
19.1 MB

**3: blastp on db**  
21,556 lines  
format: tabular, database: ?

**2: chr22\_check.gff3**

**1: human\_protein.fas**

1	2	3	4	5	6
sp P31946 1433B_HUMAN	sp A4K2U9 1433B_PONAB	100.00	246	0	0
RYLSEVASGDNKQTTVNSQQAQYQEAFAEISKKEMQPTHPIRLGLALNFSVFYIEILNSF					
DNKQTTVNSQQAQYQEAFAEISKKEMQPTHPIRLGLALNFSVFYIEILNSPEKACSLAKT					
sp P31946 1433B_HUMAN	sp Q4R572 1433B_MACFA	100.00	246	0	0
RYLSEVASGDNKQTTVNSQQAQYQEAFAEISKKEMQPTHPIRLGLALNFSVFYIEILNSF					
DNKQTTVNSQQAQYQEAFAEISKKEMQPTHPIRLGLALNFSVFYIEILNSPEKACSLAKT					



## Modification des attributs

- Renommer
- Annoter...

Galaxy / ABiMS
Analyze Data | Workflow | Shared Data | Visualization | Admin | Help | User
Using 18.0 GB

Attributes
Convert Format
Datatype
Permissions

**Edit Attributes**

**Name:**

**Info:**

**Annotation / Notes:**

blastp vs swissprot

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

**Database/Build:**

unspecified (?)

**Number of comment lines:**

This will inspect the dataset and attempt to correct the above column values if they are not accurate.

**History**

**Human chr22 proteome study**  
 19.1 MB

**3: blastp on db**  
 21,556 lines  
 format: tabular, database: ?

1	2	3	4	5	6
sp P31946 1433B_HUMAN		sp A4K2U9 1433B_PONAB	100.00	246	
RYLSEVASGDNKQTTVNSQAYQEAFEISKKEMQPTHP	IRLGLALNFSVFYIEILNSF				
DNKQTTVNSQAYQEAFEISKKEMQPTHP	IRLGLALNFSVFYIEILNSPEKACSLAKT				
sp P31946 1433B_HUMAN		sp Q4R572 1433B_MACFA	100.00	246	
RYLSEVASGDNKQTTVNSQAYQEAFEISKKEMQPTHP	IRLGLALNFSVFYIEILNSF				
DNKQTTVNSQAYQEAFEISKKEMQPTHP	IRLGLALNFSVFYIEILNSPEKACSLAKT				

**2: chr22\_check.gff3**

**1: human\_protein.fas**

81 /

# Ajout de tags et d'annotations

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 18.0 GB

Attributes Convert Format Datatype Permissions

### Edit Attributes

**Name:**

**Info:**

**Annotation / Notes:**

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

**Database/Build:**

**Number of comment lines:**

This will inspect the dataset and attempt to correct the above column values if they are not accurate.

---

**History**

**Human chr22 proteome study**  
 19.1 MB

**3: blastp\_vs\_swissprot**  
 21,556 lines  
 format: tabular, database: ?  
 Tags: blast  
 Annotation: blastp vs swissprot

1	2	3	4	5	6
sp P31946 1433B_HUMAN		sp A4K2U9 1433B_PONAB	100.00	246	
RYLSEVASGDNKQTTVNSQAYQEAFEISKKEMQPTHP		IRLGLALNFSVFYYEILNSF			
DNKQTTVNSQAYQEAFEISKKEMQPTHP		IRLGLALNFSVFYYEILNSPEKACSLAKT			
sp P31946 1433B_HUMAN		sp Q4R572 1433B_MACFA	100.00	246	
RYLSEVASGDNKQTTVNSQAYQEAFEISKKEMQPTHP		IRLGLALNFSVFYYEILNSF			
DNKQTTVNSQAYQEAFEISKKEMQPTHP		IRLGLALNFSVFYYEILNSPEKACSLAKT			

**2: chr22\_check.gff3**

**1: human\_protein.fas**

# Modification du type de fichier

- Tabular → gtf
- fastq → fastqsanger (phred33)

Galaxy / ABiMS
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 18.0 GB

Attributes
Convert Format
Datatype
Permissions

### Change data type

**New Type:**

tabular

|

rdata  
rgb  
sam  
scf  
sff  
summary\_tree  
svg  
tabix  
tabular

dataset but *not* modify its contents. Use this if Galaxy has incorrectly guessed the type of your dataset.

#### History

**Human chr22 proteome study**  
19.1 MB

**3: blastp on db**  
21,556 lines  
format: tabular, database: ?

1	2	3	4	5	6
sp P31946 1433B_HUMAN	sp A4K2U9 1433B_PONAB	100.00	246		
RYLSEVASGDNKQTTVSNSQQAYQEAFEISKKEMQPTHPIRLGLALNFSVFYYEILNSF					
DNKQTTVSNSQQAYQEAFEISKKEMQPTHPIRLGLALNFSVFYYEILNSPEKACSLAKT					
sp P31946 1433B_HUMAN	sp Q4R572 1433B_MACFA	100.00	246		
RYLSEVASGDNKQTTVSNSQQAYQEAFEISKKEMQPTHPIRLGLALNFSVFYYEILNSF					
DNKQTTVSNSQQAYQEAFEISKKEMQPTHPIRLGLALNFSVFYYEILNSPEKACSLAKT					

**2: chr22\_check.gff3**

**1: human\_protein.fas**

# Conversion de type

Dans la limite des stocks disponibles

Galaxy / ABiMS

[Analyze Data](#)
[Workflow](#)
[Shared Data](#)
[Visualization](#)
[Admin](#)
[Help](#)
[User](#)
Using 18.0 GB

Attributes
Convert Format
Datatype
Permissions

Convert to new format

Convert GFF to BED

This will create a new dataset with the contents of this dataset converted to a new format.

History
↻ ⚙

**Human chr22 proteome study**

19.1 MB 📎 📄

**3: blastp\_vs\_swissprot** 👁 ✎ ✕

**2: chr22\_check.gff3** 👁 ✎ ✕

66,141 lines, 1 comments  
 format: gff, database: ?  
 uploaded gff file 📎 📄

1. Seqname	2. Source	3. Feature	4. Start	5. End
#gff-version 3				
chr22	processed_transcript	gene	42545877	42551897
chr22	processed_transcript	transcript	42548208	42550874
chr22	processed_transcript	exon	42548208	42548246
chr22	processed_transcript	exon	42549039	42549216
chr22	processed_transcript	exon	42550831	42550874

**1: human\_protein.fas** 👁 ✎ ✕

84 /

## Visualisation

- Tabular → graphique de type scatterplot
- gff/bam → visualisation dans un système de track

Galaxy / ABiMS
Analyze Data | Workflow | Shared Data | Visualization | Admin | Help | User
Using 18.0 GB

### Scatterplot of 'blastp\_vs\_swissprot'

Data Controls

Chart Controls

Statistics

Chart

#### History

**Human chr22 proteome study**  
19.1 MB

**3: blastp\_vs\_swissprot**  
21,556 lines  
format: tabular, database: ?

1	2	3	4	5	6
sp P31946 1433B_HUMAN	sp A4K2U9 1433B_PONAB	100.00	246		
RYLSEVASGDNKQTTVSNSQQAYQEAFEISKKEMQPTHPIRLGLALNFSVFYIEILNSF					
DNKQTTVSNSQQAYQEAFEISKKEMQPTHPIRLGLALNFSVFYIEILNSPEKACSLAKT					
sp P31946 1433B_HUMAN	sp Q4R572 1433B_MACFA	100.00	246		
RYLSEVASGDNKQTTVSNSQQAYQEAFEISKKEMQPTHPIRLGLALNFSVFYIEILNSF					
DNKQTTVSNSQQAYQEAFEISKKEMQPTHPIRLGLALNFSVFYIEILNSPEKACSLAKT					

**2: chr22\_check.gff3**

**1: human\_protein.fas**

85 /

## Re-ouverture du formulaire à l'origine de l'objet avec ces options pré-remplies

Galaxy / ABiMS
Analyze Data Workflow Shared Data Visualization Admin Help User
Using 18.0 GB

### NCBI BLAST+ blastp (version 0.0.17)

**Protein query sequence(s):**  
1: human\_protein.fas

**Subject database/sequences:**  
BLAST Database

**Protein BLAST database:**  
uniprot\_swissprot

**Type of BLAST:**  
 blastp  
 blastp-short

**Set expectation value cutoff:**  
0.0001

**Output format:**  
Tabular (extended 24 columns)

**Advanced Options:**  
Hide Advanced Options

Execute

**Note.** Database searches may take a substantial amount of time. For large input datasets it is advisable to allow overnight processing.

---

**What it does**  
Search a *protein database* using a *protein query*, using the NCBI BLAST+ blastp command line tool.

**!** You can also search against a FASTA file of subject protein sequences. This is *not* advised because it is slower (only one CPU is used), but more importantly gives e-values for pairwise searches (very small e-values which will look overly significant). In most cases you should instead turn the other FASTA file into a database first using *makeblastdb* and search against that.

---

**Output format**  
Because Galaxy focuses on processing tabular data, the default output of this tool is tabular. The standard BLAST+ tabular output contains 12 columns:

Column	NCBI name	Description
1	qseqid	Query Seq-id (ID of your sequence)

**History**

**Human chr22 proteome study**  
19.1 MB

**3: blastp\_vs\_swissprot**  
21,556 lines  
format: tabular, database: ?

Run this job again

	3	4	5	6
sp P31946 1433B_HUMAN	sp A4K2U9 1433B_PONAB	100.00	246	
RYLSEVASGDNKQTTVSNSQQAYQEAFEISKKEMQPTHPIRLGLALNFSVFYIEILNSF				
DNKQTTVSNSQQAYQEAFEISKKEMQPTHPIRLGLALNFSVFYIEILNSPEKACSLAKT				
sp P31946 1433B_HUMAN	sp Q4R572 1433B_MACFA	100.00	246	
RYLSEVASGDNKQTTVSNSQQAYQEAFEISKKEMQPTHPIRLGLALNFSVFYIEILNSF				
DNKQTTVSNSQQAYQEAFEISKKEMQPTHPIRLGLALNFSVFYIEILNSPEKACSLAKT				

**2: chr22\_check.gff3**

**1: human\_protein.fas**

86 /

## Suppression d'un objet

Galaxy / ABiMS
Using 18.0 GB

Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User

### Tool: Upload File

Name:	chr22_check.gff3
Created:	Aug 09, 2013
Filesize:	5.2 MB
Dbkey:	?
Format:	gff
Galaxy Tool Version:	1.1.3
Tool Version:	
Tool Standard Output:	<a href="#">stdout</a>
Tool Standard Error:	<a href="#">stderr</a>
Tool Exit Code:	0
API ID:	e3c6542f4e108974
Full Path:	/w/galaxy/dev/galaxy-dist/database/files/001/dataset_1672.dat

Input Parameter	Value	Note for rerun
File Format	auto	
async_datasets	1984	
Specify Files for Dataset (auto)	1 uploaded datasets	
Genome	unspecified (?)	
File Format	auto	

### Inheritance Chain

chr22\_check.gff3

### History

**Human chr22 proteome study**  
 19.1 MB

**3: blastp\_vs\_swissprot**

**1: human\_protein.fas**

## Suppression d'un objet

Mais **attention**, l'objet est dans une sorte de corbeille

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 18.0 GB

### Tool: Upload File

Name:	chr22_check.gff3
Created:	Aug 09, 2013
Filesize:	5.2 MB
Dbkey:	?
Format:	gff
Galaxy Tool Version:	1.1.3
Tool Version:	
Tool Standard Output:	<a href="#">stdout</a>
Tool Standard Error:	<a href="#">stderr</a>
Tool Exit Code:	0
API ID:	e3c6542f4e108974
Full Path:	/w/galaxy/dev/galaxy-dist/database/files/001/dataset_1672.dat

Input Parameter	Value	Note for rerun
File Format	auto	
async_datasets	1984	
Specify Files for Dataset (auto)	1 uploaded datasets	
Genome	unspecified (?)	
File Format	auto	

### Inheritance Chain

chr22\_check.gff3

### History

Human chr22 proteome stu  
19.1 MB

3: [blastp\\_vs\\_swissprot](#)

**⚠ This dataset has been undelete it or [here](#) to disk**

2: [chr22\\_check.gff3](#)

1: [human\\_protein.fas](#)

**HISTORY LISTS**

- Saved Histories
- Histories Shared with Me

**CURRENT HISTORY**

- Create New
- Copy History
- Copy Datasets
- Share or Publish
- Extract Workflow
- Dataset Security
- Resume Paused Jobs
- Collapse Expanded Datasets
- Include Deleted Datasets
- Include Hidden Datasets
- Unhide Hidden Datasets
- Delete Hidden Datasets
- Purge Deleted Datasets
- Show Structure
- Export to File
- Delete
- Delete Permanently

**OTHER ACTIONS**

- Import from File



## Suppression d'un objet

Mais **attention**, l'objet est dans une sorte de corbeille  
 « Du coup, l'espace disque n'est pas libéré »

Galaxy / ABiMS
Analyze Data Workflow Shared Data Visualization Admin Help User
Using 18.0 GB

### Tool: Upload File

Name:	chr22_check.gff3
Created:	Aug 09, 2013
Filesize:	5.2 MB
Dbkey:	?
Format:	gff
Galaxy Tool Version:	1.1.3
Tool Version:	
Tool Standard Output:	<a href="#">stdout</a>
Tool Standard Error:	<a href="#">stderr</a>
Tool Exit Code:	0
API ID:	e3c6542f4e108974
Full Path:	/w/galaxy/dev/galaxy-dist/database/files/001/dataset_1672.dat

Input Parameter	Value	Note for rerun
File Format	auto	
async_datasets	1984	
Specify Files for Dataset (auto)	1 uploaded datasets	
Genome	unspecified (?)	
File Format	auto	

### Inheritance Chain

chr22\_check.gff3

### History

Human chr22 proteome stu  
19.1 MB

3: [blastp\\_vs\\_swissprot](#)

**⚠ This dataset has been undelete it or [here to disk](#)**

2: [chr22\\_check.gff3](#)

1: [human\\_protein.fas](#)

- HISTORY LISTS
- Saved Histories
- Histories Shared with Me
- CURRENT HISTORY
- Create New
- Copy History
- Copy Datasets
- Share or Publish
- Extract Workflow
- Dataset Security
- Resume Paused Jobs
- Collapse Expanded Datasets
- Include Deleted Datasets
- Include Hidden Datasets
- Unhide Hidden Datasets
- Delete Hidden Datasets
- Purge Deleted Datasets
- Show Structure
- Export to File
- Delete
- Delete Permanently
- OTHER ACTIONS
- Import from File

javascript:void(0);
89 /
>

Purge des objets deletés

Mais **attention**, là, c'est pour de bon !

Galaxy / ABiMS
Analyze Data Workflow Shared Data Visualization Admin Help User
Using 18.0 GB

### Tool: Upload File

Name:	chr22_check.gff3
Created:	Aug 09, 2013
Filesize:	5.2 MB
Dbkey:	?
Format:	gff
Galaxy Tool Version:	1.1.3
Tool Version:	
Tool Standard Output:	<a href="#">stdout</a>
Tool Standard Error:	<a href="#">stderr</a>
Tool Exit Code:	0
API ID:	e3c6542f4e108974
Full Path:	/w/galaxy/dev/galaxy-dist/database/files/001/dataset_1672.dat

Input Parameter	Value	Note for rerun
File Format	auto	
async_datasets	1984	
Specify Files for Dataset (auto)	1 uploaded datasets	
Genome	unspecified (?)	
File Format	auto	

### Inheritance Chain

chr22\_check.gff3

### History

Human chr22 proteome stu  
19.1 MB

3: blastp\_vs\_swissprot

This dataset has been undelete it or [here to disk](#)

2: chr22\_check.gff3

1: human\_protein.fas

- HISTORY LISTS
- Saved Histories
- Histories Shared with Me
- CURRENT HISTORY
- Create New
- Copy History
- Copy Datasets
- Share or Publish
- Extract Workflow
- Dataset Security
- Resume Paused Jobs
- Collapse Expanded Datasets
- Include Deleted Datasets
- Include Hidden Datasets
- Unhide Hidden Datasets
- Delete Hidden Datasets
- Purge Deleted Datasets
- Show Structure
- Export to File
- Delete
- Delete Permanently
- OTHER ACTIONS
- Import from File

- Cycle de vie des données
  - <http://abims.sb-roscoff.fr/galaxyproject>
  - Gestion par l'utilisateur de son quota
  - Suppression automatique en cours de réflexion
    - 6 mois ou 1 an

## Saved history :

- Renommer
- Supprimer [définitivement]

**Galaxy / ABiMS**

Analyze Data | Workflow | Shared Data | Visualization | Help | User

Using 41%

Tools

search tools

Get Data

ABiMS WORKFLOWS

[Workflow RNA-seq de novo by ABiMS](#)

[Workflow RNA-seq with reference by ABiMS](#)

[Workflow 4 Metabolomics](#)

ABiMS TOOLS

[Primer](#)

[RNASeq](#)

[InterEsil](#)

[Statistics](#)

[Utils](#)

[Phylogenetics](#)

[Debug](#)

COMMON TOOLS

[Text Manipulation](#)

[FASTA manipulation](#)

[Join, Subtract and Group](#)

[Filter and Sort](#)

[NCBI BLAST+](#)

[NGS: QC and manipulation](#)

[NGS: RNA Analysis](#)

[NGS: Mapping](#)

[NGS: Picard \(beta\)](#)

[NGS: SAM Tools](#)

[SVDetect](#)

[VarScan](#)

Workflows

- All workflows

## Saved Histories

search history names and tags

[Advanced Search](#)

<input type="checkbox"/>	Name	Datasets	Tags	Sharing	Size on Disk	Created	Last Updated ↑	Status
<input type="checkbox"/>	Human chr22 proteome study	2	1 Tag		20.0 MB	6 days ago	less than a minute ago	current history
<input type="checkbox"/>	X-Files	6	0 Tags		353.6 MB	Mar 25, 2013	2 minutes ago	
<input type="checkbox"/>	RNAseq de-novo	9	0 Tags		870.9 MB	Mar 20, 2013	Jun 10, 2013	
<input type="checkbox"/>	XCMS input test	1	0 Tags		130.6 MB	Apr 02, 2013	Apr 15, 2013	
<input type="checkbox"/>	RNAseq de-novo inputs	4	0 Tags		439.6 MB	Mar 20, 2013	Mar 20, 2013	

For 0 selected histories: [Rename](#) [Delete](#) [Delete Permanently](#) [Undelete](#)

Histories that have been deleted for more than a time period specified by the Galaxy administrator(s) may be permanently deleted.

History

- HISTORY LISTS
  - Saved Histories
  - Histories Shared with Me
- CURRENT HISTORY
  - Create New
  - Copy History
  - Copy Datasets
  - Share or Publish
  - Extract Workflow
  - Dataset Security
  - Resume Paused Jobs
  - Collapse Expanded Datasets
  - Include Deleted Datasets
  - Include Hidden Datasets
  - Unhide Hidden Datasets
  - Purge Deleted Datasets
  - Show Structure
  - Export to File
  - Delete
  - Delete Permanently
- OTHER ACTIONS
  - Import from File

## Saved history :

- Renommer
- Supprimer [définitivement]

Galaxy / ABiMS
Analyze Data Workflow Shared Data Visualization Help User
Using 41%

Tools

Get Data

ABiMS WORKFLOWS

[Workflow RNA-seq de novo by ABiMS](#)

[Workflow RNA-seq with reference by ABiMS](#)

[Workflow 4 Metabolomics](#)

ABiMS TOOLS

[Primer](#)

[RNASeq](#)

[InterEsil](#)

[Statistics](#)

[Utils](#)

[Phylogenetics](#)

[Debug](#)

COMMON TOOLS

[Text Manipulation](#)

[FASTA manipulation](#)

[Join, Subtract and Group](#)

[Filter and Sort](#)

[NCBI BLAST+](#)

[NGS: QC and manipulation](#)

[NGS: RNA Analysis](#)

[NGS: Mapping](#)

[NGS: Picard \(beta\)](#)

[NGS: SAM Tools](#)

[SVDetect](#)

[VarScan](#)

Workflows

- All workflows

### Saved Histories

[Advanced Search](#)

<input type="checkbox"/>	Name	Datasets	Tags	Sharing	Size on Disk	Created	Last Updated ↑	Status
<input type="checkbox"/>	Human chr22 proteome study	2	2 Tags		20.0 MB	6 days ago	14 minutes ago	
<input type="checkbox"/>	X-Files	6	0 Tags		353.6 MB	Mar 25, 2013	16 minutes ago	
<input type="checkbox"/>	RNAseq de-novo	9	0 Tags		870.9 MB	Mar 20, 2013	Jun 10, 2013	current history
<input type="checkbox"/>	XCMS input test	1	0 Tags		130.6 MB	Apr 02, 2013	Apr 15, 2013	
<input type="checkbox"/>	RNAseq de-novo inputs	4	0 Tags		439.6 MB	Mar 20, 2013	Mar 20, 2013	

For 0 selected histories: [Rename](#) [Delete](#) [Delete Permanently](#) [Undelete](#)

Histories that have been deleted for more than a time period specified by the Galaxy administrator(s) may be permanently deleted.

### History

RNAseq de-novo

870.9 MB

100: Trinity on data 98 and data 97: Assembled Transcripts

92: Dark.sample.paired.2.fastq\_good.fastq\_cutadapt.fastq\_good.fastq.nonrrna.fastq\_paired.fastq

91: Dark.sample.paired.1.fastq\_good.fastq\_cutadapt.fastq\_good.fastq.nonrrna.fastq\_paired.fastq

88: BlueLight.sample.paired.2.fastq\_good.fastq\_cutadapt.fastq\_good.fastq.nonrrna.fastq\_paired.fastq

87: BlueLight.sample.paired.1.fastq\_good.fastq\_cutadapt.fastq\_good.fastq.nonrrna.fastq\_paired.fastq

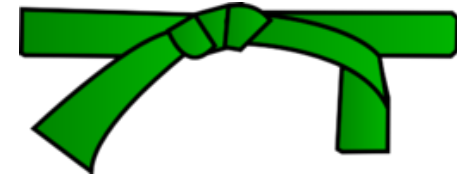
4: Dark.sample.paired.2.fastq

3: Dark.sample.paired.1.fastq

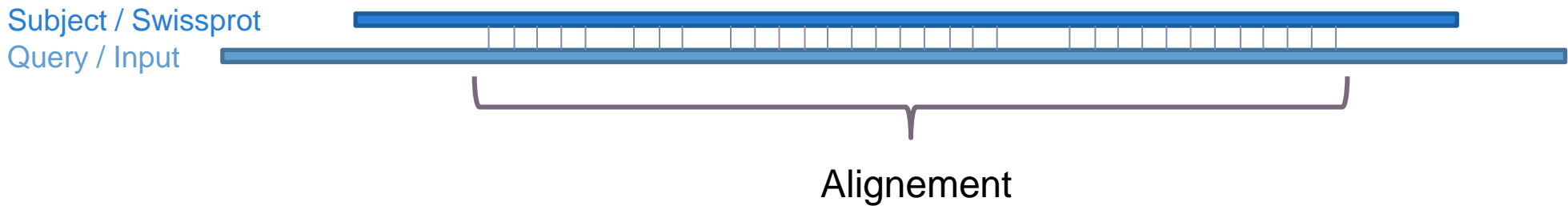
2: BlueLight.sample.paired.2.fastq

1: BlueLight.sample.paired.1.fastq

# OUTILS / EXERCICE 1



- Exercice 3<sup>e</sup> kyu :
  - Récupérer les identifiants des hits obtenus à partir de vos séquences « blastées » sur swissprot :
    - Identité > 75 %
    - Couverture de l'alignement > 75 %



$$\text{Couverture} = \text{Longueur alignement} / \text{Longueur query}$$

- Exercice 4<sup>e</sup> kyu :
  - Lancer un blast de vos séquences sur swissprot
    - Regarder quel blast convient à vos données d'entrée et à votre banque
  - Calculer le pourcentage de couverture de l'alignement par rapport à la longueur effective de la séquence query (pensez à regarder la signification des colonnes dans l'aide de blast)
  - Filtrer
    - Identité > 75 %
    - Couverture de l'alignement > 75 %







- Exercice 5<sup>e</sup> kyu :
  - Lancer un **blastx** de vos séquences sur swissprot
  - Calculer les longueurs de vos séquences
  - Fusionner le tableau de résultats du blast avec le résultat des longueurs de séquence (données tabulées)
  - Calculer le pourcentage de couverture de l'alignement par rapport à la longueur effective de la sequence query (pensez à regarder la signification des colonnes dans l'aide du blast)
  - Filtrer
    - Identité > 75 %
    - Couverture de l'alignement > 75 %



- Exercice 6<sup>e</sup> kyu :
  - Lancer un **blastx** de vos séquences sur swissprot
  - Calculer les longueurs de vos séquences (**Compute sequence length**)
  - Fusionner le tableau de résultats du blast avec le résultat des longueurs de séquence (données tabulées)
    - Si la fusion ne fonctionne pas, c'est qu'il y a un piège avec les identifiants.
  - Calculer le pourcentage de couverture de l'alignement par rapport à la longueur effective de la sequence query (pensez à regarder la signification des colonnes dans l'aide du blast)
  - Filtrer
    - Identité > 75 %
    - Couverture de l'alignement > 75 %



- Exercice 7<sup>e</sup> kyu :
  - Lancer un **blastx** de vos séquences sur swissprot
  - Calculer les longueurs de vos séquences (**Compute sequence length**)
  - Fusionner le tableau de résultats du blast avec le résultat des longueurs de séquence (données tabulées)
    - Si la fusion ne fonctionne pas, c'est qu'il y a un piège avec les identifiants.
    - Une des pistes est de travailler le tableau des longueurs de séquences pour éliminer "path ...."
  - Calculer le pourcentage de couverture de l'alignement par rapport à la longueur effective de la sequence query (pensez à regarder la signification des colonnes dans l'aide du blast)
  - Filtrer
    - Identité > 75 %
    - Couverture de l'alignement > 75 %



- Exercice 8<sup>e</sup> kyu :
  - Lancer un **blastx** de vos séquences sur swissprot
  - Calculer les longueurs de vos séquences (**Compute sequence length**)
  - Fusionner (**join**) le tableau de résultats du blast avec le résultat des longueurs de séquence (données tabulées)
    - Si la fusion ne fonctionne pas, c'est qu'il y a un piège avec les identifiants.
    - Une des pistes est de travailler le tableau des longueurs de séquences pour éliminer "path ...." en utilisant la commande **cut** et en jouant sur le délimiteur de champs. Vous pouvez ainsi obtenir 2 datasets que vous *collez* l'un à côté de l'autre
  - Calculer (**compute**) le pourcentage de couverture de l'alignement par rapport à la longueur effective de la sequence query (pensez à regarder la signification des colonnes dans l'aide du blast)
  - Filtrer (**filter**)
    - Identité > 75 %
    - Couverture de l'alignement > 75 %



- Exercice 9<sup>e</sup> kyu :
  - Lancer un **blastx** de vos séquences sur swissprot
  - Calculer les longueurs de vos séquences (**Compute sequence length**)
  - Fusionner (**join**) le tableau de résultats du blast avec le résultat des longueurs de séquence (données tabulées)
    - Si la fusion ne fonctionne pas, c'est qu'il y a un piège avec les identifiants.
    - Une des pistes est de travailler le tableau des longueurs de séquences pour éliminer "path ...."
      - Un **cut** pour récupérer la première colonne en utilisant l'espace
      - Un **cut** pour récupérer la seconde colonne en utilisant la tabulation
      - Un **paste** pour coller les 2 tableaux
  - Calculer (**compute**) le pourcentage de couverture de l'alignement par rapport à la longueur effective de la séquence query :  $((c8 - c7 + 1) / c14) * 100$
  - Filtrer (**filter**)
    - $(c3 \geq 75 \text{ and } \text{abs}(c15) \geq 75)$
    - Identité > 75 % et Couverture de l'alignement > 75 %

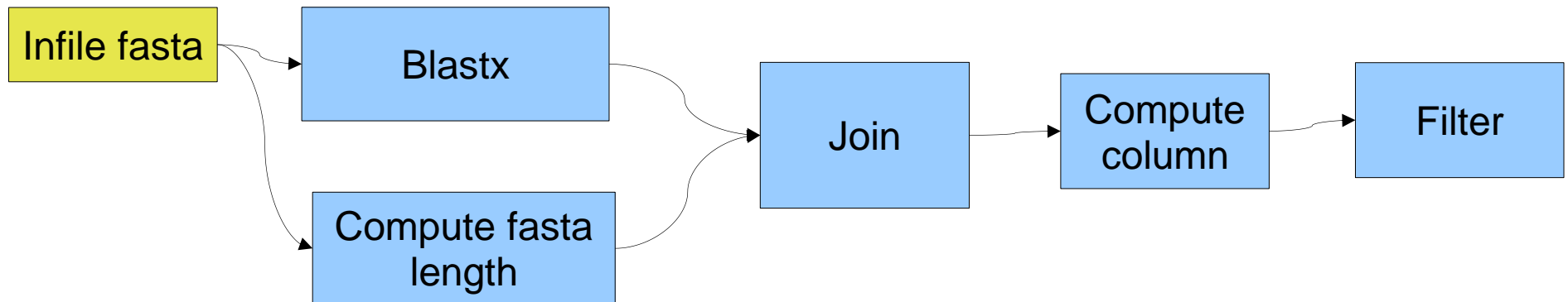
# OUTILS / EXERCICE 2

# WORKFLOW

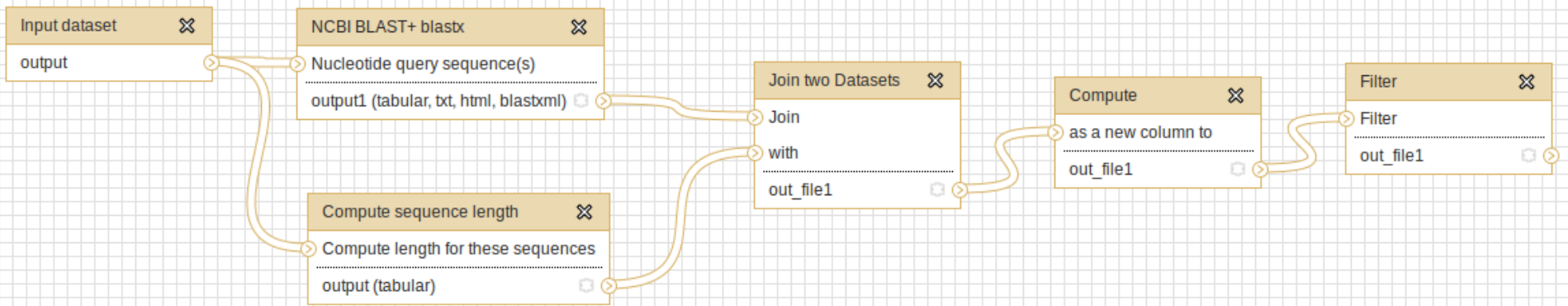
- Un workflow est un enchaînement d'outils et de paramètres
- Peut correspondre au protocole de l'expérience
- Un workflow est construit pour être rejoué (de manière plus ou moins stricte)



- Notre workflow



# Notre workflow sauce Galaxy



106 /

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 18.1 GB

Tools search tools

**Get Data**

ABiMS WORKFLOWS

[Workflow RNA-seq de novo by ABiMS](#)

[Workflow RNA-seq with reference by ABiMS](#)

[Workflow 4 Metabolomics](#)

ABiMS TOOLS

[Primer](#)

[RNASeq](#)

[InterEsil](#)

[Statistics](#)

[Utils](#)

[Phylogenetics](#)

[Debug](#)

COMMON TOOLS

[Text Manipulation](#)

[FASTA manipulation](#)

[Join, Subtract and Group](#)

[Filter and Sort](#)

[Graphics](#)

[NCBI BLAST+](#)

[NGS: QC and manipulation](#)

[NGS: RNA Analysis](#)

[NGS: Mapping](#)

[NGS: Picard \(beta\)](#)

[NGS: SAM Tools](#)

[NGS: GATK Tools \(beta\)](#)

[SVDetect](#)

[VarScan](#)

[Muscle](#)

[RAxML](#)


**Online**

**i**

- 07-06-13: Metabolomic : Workflow 4 Metabolomics, updated to version 2.1.0 (2013\_06\_07) [i](#)
- 30-04-13: RNASeq : DESeq is now available for RNASeq expression data with reference (with gtf input).
- 26-04-13: RNASeq : DESeq is now available for denovo RNASeq expression data (without gtf input).
- 26-04-13: RNASeq : sam2counts is now available to count the reads coverage by transcrit. It's also a requirement for DESeq denovo.
- 26-04-13: Metabolomic : Workflow Metabolomic by ABiMS, updated to version 2.0.0 (2013\_04\_18) [i](#)

**Ab4iMS**

**Analyses and Bioinformatics for Marine Science**

 CNRS UPMC  
**Station Biologique Roscoff**

**i** **Information**  
For any question or request for tools or account, send an email at [support.abims 'AT' sb-roscoff.fr](mailto:support.abims@at.sb-roscoff.fr)

Galaxy is an open, web-based platform for data intensive biomedical research. The [Galaxy team](#) is a part of [BX](#) at [Penn State](#), and the [Biology and Mathematics and Computer Science](#) departments at [Emory University](#). The [Galaxy Project](#) is supported in part by [NHGRI](#), [NSF](#), [The Huck Institutes of the Life Sciences](#), [The Institute for CyberScience at Penn State](#), and [Emory University](#).

**History** refresh gear

**HISTORY LISTS**

Annotation  
2.8 MB

10: blastx

8: blastx

7: blastx

6: fast

4: Joh

2: blas

1: Joh

Saved Histories

Histories Shared with Me

**CURRENT HISTORY**

Create New

Copy History

Copy Datasets

Share or Publish

**Extract Workflow**

Dataset Security

Resume Paused Jobs

Collapse Expanded Datasets

Include Deleted Datasets

Include Hidden Datasets

Unhide Hidden Datasets

Delete Hidden Datasets

Purge Deleted Datasets

Show Structure

Export to File

Delete

Delete Permanently

**OTHER ACTIONS**

Import from File

107 /

# Création d'un workflow à partir d'un historique

- Changement du nom
- Possibilité de décocher des étapes

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 18.1 GB

Tools search tools

Get Data

ABiMS WORKFLOWS

[Workflow RNA-seq de novo by ABiMS](#)

[Workflow RNA-seq with reference by ABiMS](#)

[Workflow 4 Metabolomics](#)

ABiMS TOOLS

[Primer](#)

[RNASeq](#)

[InterEsil](#)

[Statistics](#)

[Utils](#)

[Phylogenetics](#)

[Debug](#)

COMMON TOOLS

[Text Manipulation](#)

[FASTA manipulation](#)

[Join, Subtract and Group](#)

[Filter and Sort](#)

[Graphics](#)

[NCBI BLAST+](#)

[NGS: QC and manipulation](#)

[NGS: RNA Analysis](#)

[NGS: Mapping](#)

[NGS: Picard \(beta\)](#)

[NGS: SAM Tools](#)

[NGS: GATK Tools \(beta\)](#)

[SVDetect](#)

[VarScan](#)

[Muscle](#)

[RAxML](#)

The following list contains each tool that was run to create the datasets in your current history. Please select those that you wish to include in the workflow.

Tools which cannot be run interactively and thus cannot be incorporated into a workflow will be shown in gray.

**Workflow name**

Workflow constructed from history 'Annotation John Doe'

Create Workflow Check all Uncheck all

Tool	History items created
Upload File <i>This tool cannot be used in workflows</i>	1: JohnDoe.fasta Treat as input dataset
NCBI BLAST+ blastx <input checked="" type="checkbox"/> Include "NCBI BLAST+ blastx" in workflow	2: blastxOnSwissprot.tab
Cut <input checked="" type="checkbox"/> Include "Cut" in workflow	4: JohnDoe.ids.fasta
Compute sequence length <input checked="" type="checkbox"/> Include "Compute sequence length" in workflow	6: fastaLength.tab
Join two Datasets <input checked="" type="checkbox"/> Include "Join two Datasets" in workflow	7: blastxOnSwissprot_length.tab
Compute <input checked="" type="checkbox"/> Include "Compute" in workflow	8: blastxOnSwissprot_coverage.tab
Filter <input checked="" type="checkbox"/> Include "Filter" in workflow	10: blastxOnSwissprot_i50_c50.tab

History

Annotation John Doe  
2.8 MB

10: blastxOnSwissprot\_i50\_c50.tab

8: blastxOnSwissprot\_coverage.tab

7: blastxOnSwissprot\_length.tab

6: fastaLength.tab

4: JohnDoe.ids.fasta

2: blastxOnSwissprot.tab

1: JohnDoe.fasta

108 /

- Possibilité de créer un workflow de zéro

Galaxy / ABiMS Analyze Data **Workflow** Shared Data Visualization Admin Help User Using 18.1 GB

### Your workflows

[+ Create new workflow](#) [Upload or import workflow](#)

Name	# of Steps
<a href="#">Blast and Filter</a>	7
<a href="#">RNASeq - denovo</a>	6
<a href="#">Workflow XCMS-CAMERA-Stat</a>	14

### Workflows shared with you by others

No workflows have been shared with you.

### Other options

[Configure your workflow menu](#)

Galaxy / ABiMS Analyze Data **Workflow** Shared Data Visualization Admin Help User Using 18.1 GB

### Your workflows

[+ Create new workflow](#) [↑ Upload or import workflow](#)

Name	# of Steps
Blast and Filter	7
R	6
W	14

**Workflows shared with you by others**

No workflows shared with you.

**Other actions**

- Edit
- Run
- Share or Publish
- Download or Export
- Copy
- Rename
- View
- Delete

Configure your workflow menu

# Le canevas · Glisser-Déposé

Galaxy / ABiMS Analyze Data **Workflow** Shared Data Visualization Admin Help User Using 18.1 GB

Tools Workflow Canvas | Blast and Filter Details

search tools

**Get Data**

ABiMS WORKFLOWS

[Workflow RNA-seq de novo by ABiMS](#)

[Workflow RNA-seq with reference by ABiMS](#)

[Workflow 4 Metabolomics](#)

ABiMS TOOLS

[Primer](#)

[RNASeq](#)

[InterEsil](#)

[Statistics](#)

[Utils](#)

[Phylogenetics](#)

[Debug](#)

COMMON TOOLS

[Text Manipulation](#)

[FASTA manipulation](#)

[Join, Subtract and Group](#)

[Filter and Sort](#)

[Graphics](#)

[NCBI BLAST+](#)

[NGS: QC and manipulation](#)

[NGS: RNA Analysis](#)

[NGS: Mapping](#)

[NGS: Picard \(beta\)](#)

[NGS: SAM Tools](#)

[NGS: GATK Tools \(beta\)](#)

[SVDetect](#)

[VarScan](#)

[Muscle](#)

```

graph LR
    A[Input dataset] --> B[NCBI BLAST+ blastx]
    B --> C[Join two]
    C --> D[Cut]
    C --> E[Compute sequence length]
    D --> E
  
```

**Edit Workflow Attributes**

**Name:**  
Blast and Filter

**Tags:**

Apply tags to make it easy to search for and find items with the same tag.

**Annotation / Notes:**  
Describe or add notes to workflow  
Add an annotation or notes to a workflow; annotations are available when a workflow is viewed.

# Edition d'un workflow :

- Ajout d'un nouvel outil dans le workflow

The screenshot displays the Galaxy ABiMS interface. On the left, a sidebar lists various tools under categories like 'Text Manipulation'. The 'Cut columns from a table' tool is highlighted with a red box and a red number '1'. The main area is the 'Workflow Canvas | Blast and Filter', showing a workflow with three steps: 'Compute' (output: out\_file1), 'Filter' (output: out\_file1), and 'Cut' (output: out\_file1 (tabular)). The 'Cut' tool is highlighted with a red box and a red number '2'. On the right, a 'Details' panel for the 'Cut' tool shows its version (1.0.2), configuration (Cut columns: c1,c2; Delimited by: Tab), and a warning: 'WARNING: This tool breaks column assignments. To re-establish column assignments run the tools and click on the pencil icon in the latest history item.'



- Ajout d'un nouvel outil dans le workflow

The screenshot shows the Galaxy ABiMS interface with a workflow canvas titled "Workflow Canvas | Blast and Filter". The workflow consists of three tools connected in sequence: "Compute", "Filter", and "Cut".

- Compute:** "as a new column to" with output "out\_file1".
- Filter:** "Filter" with output "out\_file1".
- Cut:** "From" with output "out\_file1 (tabular)".

Red annotations highlight specific parts of the interface:

- 3 - save:** A red box around the gear icon in the top right corner of the workflow canvas.
- 2:** A red box around the "Cut columns:" dropdown menu in the "Details" panel, which is set to "c1".
- 1:** A red box around the "From" field in the "Cut" tool's configuration, which is set to "out\_file1 (tabular)".

The "Details" panel on the right shows the configuration for the "Cut" tool (Version: 1.0.2). It includes a "Cut columns:" dropdown set to "c1", a "Delimited by:" dropdown set to "Tab", and a "From" field set to "Data input 'input' (txt)". Below this, there are "Edit Step Actions" (Rename Dataset, out\_file1, Create) and "Edit Step Attributes" (Annotation / Notes).

A warning message is visible at the bottom right: "WARNING: This tool breaks column assignments. To re-establish column assignments run the tools and click on the pencil icon in the latest history item." Another note states: "The output of this tool is always in tabular format (e.g., if your original delimiters are commas they will be" followed by a redacted area).

- Rendre paramétrable un paramètre au moment de l'exécution

The screenshot shows the Galaxy ABiMS interface with a workflow titled "Workflow Canvas | Blast and Filter". The workflow consists of several steps: "Input dataset" (output: "output"), "NCBI BLAST+ blastx" (input: "output", output: "output1 (tabular, txt, html, blastxml)"), "Join two Datasets" (inputs: "output1" and "out\_file1", output: "Join"), "Cut" (input: "Join", output: "out\_file1 (tabular)"), and "Compute sequence length" (input: "out\_file1", output: "output (tabular)").

The "NCBI BLAST+ blastx" tool configuration is highlighted with a red box. The "Protein BLAST database:" dropdown is also highlighted with a red box and labeled with a red "2". The "Set at runtime" option is selected and labeled with a red "3". A red "1" is placed near the output of the BLAST step.

The right sidebar shows the tool configuration for "NCBI BLAST+ blastx", version 0.0.17. It includes fields for "Nucleotide query sequence(s)", "Subject database/sequences:", "Protein BLAST database:", "Query genetic code:", "Set expectation value cutoff:", and "Output format:". The "Edit Step Actions" section shows "Rename Dataset" and "output1 Create".

At the bottom left, the URL "galaxydev.sb-roscoff.fr/workflow/editor?id=1749d571c4b6b63b#" is visible. At the bottom right, the page number "114 /" is shown.

Galaxy / ABiMS Analyze Data **Workflow** Shared Data Visualization Admin Help User Using 18.1 GB

### Your workflows

[Create new workflow](#) [Upload or import workflow](#)

Name	# of Steps
Blast and Filter	7
R	6
W	14

**Workflows shared with you by others**

No workflows shared with you.

**Other actions**

- Edit
- Run
- Share or Publish
- Download or Export
- Copy
- Rename
- View
- Delete

Configure your workflow menu

- Paramètres par défaut
- Paramètres à renseigner

Galaxy / ABiMS

Analyze Data Workflow Shared Data Visualization Admin Help User Using 18.1 GB

Tools

search tools

Get Data

ABiMS WORKFLOWS

[Workflow RNA-seq de novo by ABiMS](#)

[Workflow RNA-seq with reference by ABiMS](#)

[Workflow 4 Metabolomics](#)

ABiMS TOOLS

[Primer](#)

[RNASeq](#)

[InterEsil](#)

[Statistics](#)

[Utils](#)

[Phylogenetics](#)

[Debug](#)

COMMON TOOLS

[Text Manipulation](#)

[FASTA manipulation](#)

[Join, Subtract and Group](#)

[Filter and Sort](#)

[Graphics](#)

[NCBI BLAST+](#)

[NGS: QC and manipulation](#)

[NGS: RNA Analysis](#)

[NGS: Mapping](#)

[NGS: Picard \(beta\)](#)

[NGS: SAM Tools](#)

[NGS: GATK Tools \(beta\)](#)

[SVDetect](#)

[VarScan](#)

[Muscle](#)

[RAxML](#)

### Running workflow "Blast and Filter"

Expand All Collapse

**Step 1: Input dataset**

**Fasta file**

1: Spiderman.fasta

type to filter

**Step 2: NCBI BLAST+ blastx (version 0.0.17)**

**Nucleotide query sequence(s)**  
Output dataset 'output' from step 1

**Subject database/sequences**  
BLAST Database

**Protein BLAST database**

nr

None

None

**Query genetic code**  
1. Standard

**Set expectation value cutoff**  
0.001

**Output format**  
Tabular (extended 24 columns)

**Advanced Options**  
Hide Advanced Options

**Step 3: Cut (version 1.0.2)**

**Step 4: Compute sequence length (version 1.0.0)**

**Step 5: Join two Datasets (version 2.0.2)**

**Step 6: Compute (version 1.1.0)**

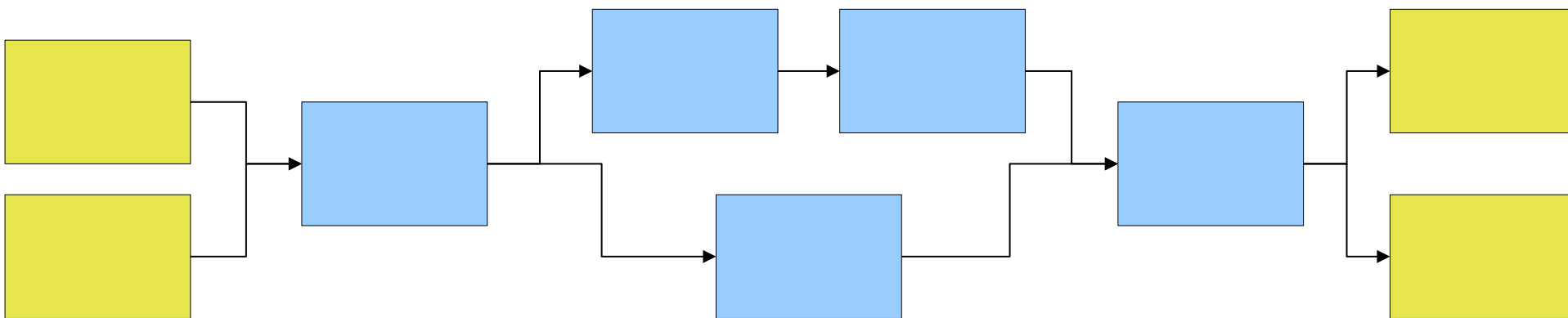
History

Unnamed history  
29.4 KB

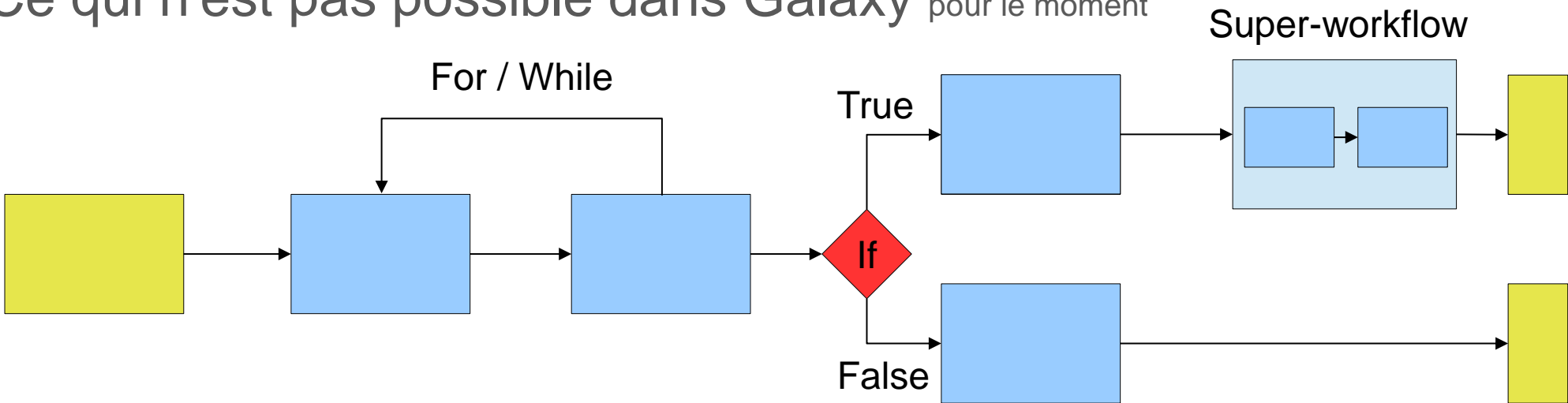
**1: Spiderman.fasta**

116 /

- Ce qui est possible dans Galaxy



- Ce qui n'est pas possible dans Galaxy pour le moment



# PARTAGE

« biologiste » ↔ « biologiste »

- Partage de datasets
  - Historique en entier
    - Avec ou sans workflow associé (Extract workflow)
- Sous ensembles

bioanalyste → « biologiste »

- Partage de workflows
  - Paramètres pré-configurés
  - Paramètres à configurer (Set at runtime)
  - => Suivant le niveau de l'utilisateur
- Partage de pages  
(voir plus loin)



bioinformaticien → bioinformaticien

- Partage de descriptions d'outils et/ou de scripts

Toolshed Galaxy

Galaxy / ABiMS

Tools

search tools

Get Data

ABiMS WORKFLOWS

- Workflow RNA-seq de novo by ABiMS
- Workflow RNA-seq with reference by ABiMS
- Workflow 4 Metabolomics

ABiMS TOOLS

- Primer
- RNASeq
- InterEsil
- Statistics
- Utils
- Phylogenetics
- Debug

COMMON TOOLS

- Text Manipulation
- FASTA manipulation
- Join, Subtract and Group
- Filter and Sort
- NCBI BLAST+
- NGS: QC and manipulation
- NGS: RNA Analysis
- NGS: Mapping
- NGS: Picard (beta)
- NGS: SAM Tools
- Muscle
- SVDetect
- VarScan
- RAxML

Workflows

### Saved Histories

search history names and tags

Advanced Search

<input type="checkbox"/>	Name	Datasets	Tags	Sharing	Size on Disk	Created	Last Updated ↑	Status
<input type="checkbox"/>	JohnDoe	4	2 Tags		57.0 MB	Aug 07, 2013	less than a minute ago	
<input type="checkbox"/>		6	0 Tags		353.6 MB	Mar 25, 2013	Aug 13, 2013	
<input type="checkbox"/>		9	0 Tags		870.9 MB	Mar 20, 2013	Jun 10, 2013	current history
<input type="checkbox"/>		1	0 Tags		130.6 MB	Apr 02, 2013	Apr 15, 2013	
<input type="checkbox"/>	RNAseq de-novo inputs	4	0 Tags		439.6 MB	Mar 20, 2013	Mar 20, 2013	

For 0 selected histories:

Histories that have been deleted for more than a time period specified by the Galaxy administrator(s) may be permanently deleted.

### History

- RNAseq de-novo  
870.9 MB
- 100: Trinity on data 98 and data 97: Assembled Transcripts
- 92: Dark.sample.paired.2.fastq\_good.fastq\_cutadapt.fastq\_good.fastq.nonrrna.fastq\_paired.fastq
- 91: Dark.sample.paired.1.fastq\_good.fastq\_cutadapt.fastq\_good.fastq.nonrrna.fastq\_paired.fastq
- 88: BlueLight.sample.paired.2.fastq\_good.fastq\_cutadapt.fastq\_good.fastq.nonrrna.fastq\_paired.fastq
- 87: BlueLight.sample.paired.1.fastq\_good.fastq\_cutadapt.fastq\_good.fastq.nonrrna.fastq\_paired.fastq
- 4: Dark.sample.paired.2.fastq
- 3: Dark.sample.paired.1.fastq
- 2: BlueLight.sample.paired.2.fastq
- 1: BlueLight.sample.paired.1.fastq

## Your workflows

[+ Create new workflow](#) [↑ Upload or import workflow](#)

Name	# of Steps
Blast and Filter	7
R	6
W	14

- Share or Publish
- Edit
- Run
- Download or Export
- Copy
- Rename
- View
- Delete

Workflows shared with you by others

## Share or Publish History 'JohnDoe'

### Make History Accessible via Link and Publish It

This history is currently restricted so that only you and the users listed below can access it. You can:

[Make History Accessible via Link](#)

Generates a web link that you can share with other people so that they can view and import the history.

→ communauté restreinte

[Make History Accessible and Publish](#)

Makes the history accessible via link (see above) and publishes the history to Galaxy's [Published Histories](#) section, w

→ tous les utilisateurs du serveur Galaxy

### Share History with Individual Users

You have not shared this history with any users.

[Share with a user](#)

→ communauté restreinte

[Back to Histories List](#)



- Import d'un published \*

Galaxy / ABiMS Analyze Data Workflow **Shared Data** Visualization Help User Using 42%

### Published Histories

search name, annotation, owner, and tags

[Advanced Search](#)

Name	Annotation	Owner	Community Tags	Last Updated
<a href="#">JohnDoe</a>		lecorguille	★★★★★ human blast	less than a minute ago

- Data Libraries
- Published Histories
- Published Workflows
- Published Visualizations
- Published Pages

- Import d'un shared history

Galaxy / ABiMS Analyze Data Workflow **Shared Data** Visualization Help User Using 42%

### Histories shared with you by others

Name	Datasets	Created	Last Updated	Shared by
<input type="checkbox"/> <a href="#">TestSPE_positivemode</a>	50	Apr 15, 2013	Jun 03, 2013	@sb-roscoff.fr

For 0 selected histories:

- HISTORY LISTS
- Saved Histories
- Histories Shared with Me**
- CURRENT HISTORY
- Create New
- Copy History
- Copy Datasets



## Level 5

- Partage d'outils et de descriptions via le toolshed

## Level 4



- Lancement des outils de manière autonome.
- Utilisation des options avancées.
- Utilisation de l'API Galaxy
- Propose des workflows aux collègues de niveau --



## Level 3

- Lancement des outils de manière autonome.



- Utilisation d'un workflow plus ou moins préconfiguré

## Level 2

- Utilisation d'un workflow préconfiguré



## Level 1

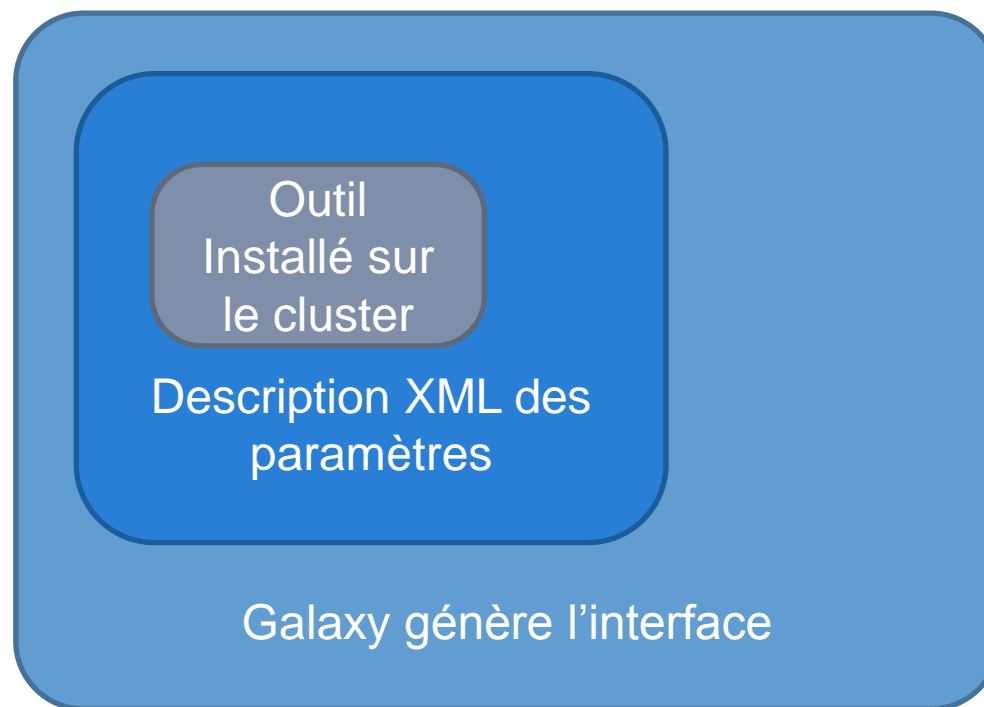
- Partage des données avec un collègue de niveau ++

# BONUS

# BONUS - MECANIQUE



- Comment un outil arrive dans Galaxy ?



## • Comment un outil arrive dans Galaxy ?

```
[lecorguille@n0 ~]$ e-PCR --help
e-PCR: invalid option -- -
usage: [-hV] [posix-options] stsfile [fasta ...]
[compat-options]
where posix-options are:
    -m ##      Margin (default 50)
    -w ##      Wordsize (default 7)
    -n ##      Max mismatches allowed (default 0)
    -g ##      Max indels allowed (default 0)
    -f ##      Use ## discontinuous words, slow
                ##>1
    -o ##      Set output file
    -t ##      Set output format:
                1 - classic, range (pos1..pos2)
                2 - classic, midpoint
                3 - tabular
                4 - tabular with alignment in c
                    (slow)
    -d##-##    Set default size range
                (default 100-350)
    -p +-      Turn hits postprocess on/off
    -v ##      Verbosity flags
    -a a|f     Use presize alignmens (only if
                gaps>0), slow
                a - Always or f - as Fallback
    -x +-      Use 5'-end lowercase masking of
                primers (default -)
    -u +-      Uppercase all primers (default -)
```

[...]

The screenshot shows the Galaxy / ABiMS interface for the e-PCR tool (version 1.0.0). The configuration page includes several input fields and dropdown menus for various parameters:

- STS file:** A dropdown menu showing "100: (as tabular) Trinity on data 9..Transcripts". Below it, the format is set to "tabular".
- Fasta file:** A dropdown menu showing "100: Trinity on data 9..Transcripts". Below it, the format is set to "fasta".
- Wordsize (W):** A text input field containing the value "7". Below it, a description states: "Set word size for primers hash (nucleotide positions). Longer word size decreases hash collision rate, but increases memory usage. Also no mismatches are allowed within word size near 'inner' boundary of primers unless one uses discontinuous words, and no gaps are ever allowed in that region."
- Use ## discontinuous words (F):** A text input field containing the value "1". Below it, a description states: "Set discontinuous word count for primers hash (1 means 'use contiguous words'). Discontinuous words increase number of hash tables and decrease 'effective' word size (thus increasing hash collision rate), so make search significantly slower, but increase sensitivity by allowing mismatches within word size. Reasonable values are 1 (contiguous words) and 3."
- Margin (M):** A text input field containing the value "50". Below it, a description states: "Set maximal allowed deviation of hit product size from expected STS size."
- Set default sts lower size (D):** A text input field containing the value "100". Below it, a description states: "Set ddefault STS size range - values used for STSs that have no size associated in file."
- Set default sts higher size (D):** A text input field containing the value "400". Below it, a description states: "Set ddefault STS size range - values used for STSs that have no size associated in file."
- Max mismatches allowed (N):** A text input field containing the value "0". Below it, a description states: "Set maximal number of mismatches allowed in primer-to-sequence alignment (per primer!)."
- Max indels allowed (G):** A text input field containing the value "0". Below it, a description states: "Set maximal number of gaps allowed in primer-to-sequence alignment (per primer!)."
- Set output format (T):** A dropdown menu showing "tabular". Below it, the text "Output formats" is visible.

At the bottom of the configuration page, there is a blue "Execute" button.

## • Comment un outil arrive dans Galaxy ?

```

<tool id="abims_epcr" name="e-PCR">
  <!-- author : lecorguille@sb-roscoff.fr -->
  <!-- date : 11-05-12 -->
  <description>e-PCR parses stsfile in unists format, then reads nucleotide sequence data in FASTA format from files listed in commandline if
  any, or from stdin otherwise. For input sequences e-PCR finds matches and prints output in one of three formats.</description>
  <command>e-PCR -w $wordsize -f $wordcnt -m $margin -d$sts_size_lo-$sts_size_hi -n $max_mismatch -g $max_gap -t $output_format $infile_stsfile
  $infile_fasta > $output</command>
  <inputs>
    <param name="infile_stsfile" type="data" label="STS file" format="tabular" help="format : tabular" />
    <param name="infile_fasta" type="data" label="Fasta file" format="fasta" help="format : fasta" />
    <param name="wordsize" type="integer" label="Wordsize (W)" value="7" help="Set word size for primers hash (nucleotide positions).
    Longer word size decreases hash collision rate, but increases memory usage. Also no mismatches are allowed within word size near
    'inner' boundary of primers unless one uses discontinuous words, and no gaps are ever allowed in that region." />
    <param name="wordcnt" type="integer" label="Use ## discontinuous words (F)" value="1" help="Set discontinuous word count for primers
    hash (1 means 'use contiguous words'). Discontinuous words increase number of hash tables and decrease 'effective' word size (thus
    increasing hash collision rate), so make search significantly slower, but increase sensitivity by allowing mismatches within word
    size. Reasonable values are 1 (contiguous words) and 3." />
    <param name="margin" type="integer" label="Margin (M)" value="50" help="Set maximal allowed deviation of hit product size from
    expected STS size." />
    <param name="sts_size_lo" type="integer" label="Set default sts lower size (D)" value="100" help="Set ddefault STS size range - values
    used for STSs that have no size associated in file." />
    <param name="sts_size_hi" type="integer" label="Set default sts higher size (D)" value="400" help="Set ddefault STS size range -
    values used for STSs that have no size associated in file." />
    <param name="max_mismatch" type="integer" label="Max mismatches allowed (N)" value="0" help="Set maximal number of mismatches allowed
    in primer-to-sequence alignment (per primer!)." />
    <param name="max_gap" type="integer" label="Max indels allowed (G)" value="0" help="Set maximal number of gaps allowed in primer-to-
    sequence alignment (per primer!)." />
    <param name="output_format" type="select" help="Output formats">
      <label>Set output format (T)</label>
      <option value="1">classic, range (pos1..pos2)</option>
      <option value="2">classic, midpoint</option>
      <option value="3" selected="true">tabular</option>
      <option value="4">tabular with alignment in comments (slow)</option>
    </param>
  </inputs>
  <outputs>
    <data name="output" format="tabular" />
  </outputs>
  <help>
  Full documentation: ftp://ftp.ncbi.nlm.nih.gov/pub/schuler/e-PCR/README.txt
  </help>

```

## • Comment un outil arrive dans Galaxy ?

```
[lecorguille@n0 ~]$ e-PCR --help
e-PCR: invalid option -- -
usage: [-hV] [posix-options] stsfile [fasta ...]
[compat-options]
where posix-options are:
  -m ##      Margin (default 50)
  -w ##      Wordsize (default 7)
  -n ##      Max mismatches allowed (default
  -g ##      Max indels allowed (default 0)
  -f ##      Use ## discontinuos words, slow
              ##>1
  -o ##      Set output file
```

```
<tool id="abims_epcr" name="e-PCR">
  <!-- author : lecorguille@sb-roscoff.fr -->
  <!-- date : 11-05-12 -->
  <description>e-PCR parses stsfile in unists format, then reads nucleotide sequence data in FASTA format from files listed in commandline if
any, or from stdin otherwise. For input sequences e-PCR finds matches and prints output in one of three formats.</description>
  <command>e-PCR -w $wordsize -f $wordcnt -m $margin -d$sts_size_lo-$sts_size_hi -n $max_mismatch -g $max_gap -t $output_format $infile_stsfile
$infile_fasta > $output</command>
  <inputs>
    <param name="infile_stsfile" type="data" label="STS file" format="tabular" help="format : tabular" />
    <param name="infile_fasta" type="data" label="Fasta file" format="fasta" help="format : fasta" />
    <param name="wordsize" type="integer" label="Wordsize (W)" value="7" help="Set word size for primers hash (nucleotide positions).
Longer word size decreases hash collision rate, but increases memory usage. Also no mismatches are allowed within word size near
'inner' boundary of primers unless one uses discontinuous words, and no gaps are ever allowed in that region." />
    <param name="wordcnt" type="integer" label="Use ## discontinuos words (F)" value="1" help="Set discontinuous word count for primers
hash (1 means 'use contiguous words'). Discontinuos words increase number of hash tables and decrease 'effective' word size (thus
increasing hash collision rate), so make search significantly slower, but increase sencitivity by allowing mismatches within word
size. Reasonable values are 1 (contiguous words) and 3." />
    <param name="margin" type="integer" label="Margin (M)" value="50" help="Set maximal allowed deviation of hit product size from
expected STS size." />
    <param name="sts_size_lo" type="integer" label="Set default sts lower size (D)" value="100" help="Set ddefault STS size range - values
used for STSs that have no size associated in file." />
    <param name="sts_size_hi" type="integer" label="Set default sts higher size (D)" value="400" help="Set ddefault STS size range -
values used for STSs that have no size associated in file." />
    <param name="max_mismatch" type="integer" label="Max mismatches allowed (N)" value="0" help="Set maximal number of mismatches allowed
in primer-to-sequence alignment (per primer!)." />
  </inputs>
</tool>
```

Galaxy / ABiMS

e-PCR (version 1.0.0)

STS file:  
100: (as tabular) Trinity on data 9..Transcripts  
format : tabular

Fasta file:  
100: Trinity on data 9..Transcripts  
format : fasta

Wordsize (W):  
7

Set word size for primers hash (nucleotide positions). Longer word size decreases hash collision rate, but increases memory usage. Also no mismatches are allowed within word size near 'inner' boundary of primers unless one uses discontinuous words, and no gaps are ever allowed in that region.

Use ## discontinuos words (F):  
1

# CONCLUSION

- Comment un outil arrive dans Galaxy ?
  - Tooshed
  - Description maison
- Une seul adresse :  
=> [support.abims@sb-roscoff.fr](mailto:support.abims@sb-roscoff.fr)