



Biogenouest
BIOGÉNOUEST



ifb
INSTITUT FRANÇAIS
DE BIOINFORMATIQUE



EMBRC
FRANCE
CENTRE NATIONAL DE RESSOURCES BIOLOGIQUES MARINES



OCEANOMICS



IDEALG
seaweed for the future

AB⁴BIMS

16/05/2019



Galaxy

Initiation

Lorraine Guéguen

Erwan Corre

Credits to Gildas Le Corguillé, Galaxy Training Network

v2.5



SORBONNE
UNIVERSITÉ
CRÉATEURS DE FUTURS
DEPUIS 1257



- Slides and datasets available:
<https://tinyurl.com/training-galaxy-initiation>
- <http://galaxy.sb-roscoff.fr>
- Login:
 - login@sb-roscoff.fr
 - *****

- Learning objectives:
 - Familiarize yourself with the basics of Galaxy
 - Learn how to import data
 - Learn how to run tools
 - Learn how histories work
 - Learn how to create a workflow
 - Learn how to share your work
 - Understand and master dataset collections

- Introduction
- Data import
- Tools
- Dataset
- History
- Workflow
- Share

- Introduction
- Data import
- Tools
- Dataset
- History
- Workflow
- Share



Sequence files manipulation



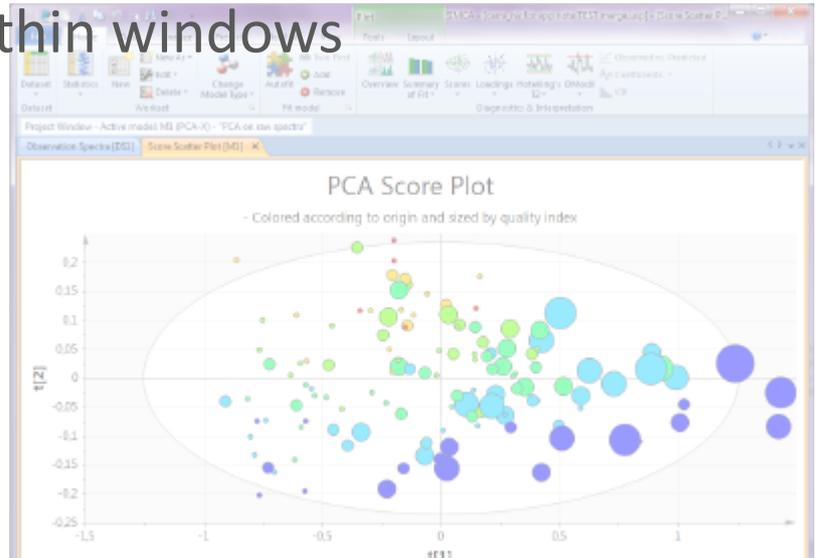
INTRODUCTION / PROBLEMATIC

- In biomedical research, high-throughput technologies produce **large datasets**.
- How to perform analyses of these data **without bioinformatics skills** ?
 - Assemble transcript sequences de novo
 - Determine the gene expression
 - Build a phylogenetic tree
 - Predict subcellular targeting for proteins
 - Identify and quantify metabolites detected by LC-MS
 - ...

Introduction / problematic

- **Graphical interface click-button tools within windows**

- + very ergonomic
- too ergonomic → lack of flexibility
- few
- paying for it!



- **Tools available on the internet**

- + very ergonomic
- too ergonomic → lack of flexibility
- A small part of the available tools
- the submission size /storage is often limited
- must not be paranoid

MetaboAnalyst 2.0
-- a comprehensive tool suite

Processing MS spectra :
The (GC/LC)-MS spectra processing is performed sequentially from left to right: Raw Data, Pre-Processing, Peak Detection, Peak List, and Peak Annotation.

Option	Value	Default	Description
use	Yes	Yes	Use the selected method for peak detection.
minimum peak width	1	1	Minimum peak width in minutes.
maximum peak width	10	10	Maximum peak width in minutes.
Signal-to-noise threshold	10	10	Signal-to-noise threshold for peak detection.
smoothing	500	500	Smoothing window size in minutes.
degree of polynomial	1	1	Degree of polynomial for smoothing.
peak list format	CSV	CSV	Format for the peak list output.
peak list columns	5	5	Number of columns for the peak list output.
peak list header	1000	1000	Number of rows for the peak list header.

XCMS
Metabolic Profile
Peak List

Introduction / problematic

```
library(xcms)
loaddata()
polar = "Polar"

noise=250000
xset <- xcmsSet(cdfFiles, ppm=ppm, mzdiff=mzwid, peakwidth=peakwidth, noise=noise, snthresh=snth, method="centWave", fitgauss=TRUE, nSlaves=8)
xset2 <- retcor(xset)
dev.copy2pdf(device = 2, file = paste(pathResult, "/Ret_Cor-Graph", expe, "_", polar, ".pdf", sep=""), paper="a4", height=9, width=14)
xset3 <- group(xset2, minfrac = 0.2, bw = bw, minsamp = 1, mzwid = mzwid, max = 50, sleep = 0)
xset5 <- fillPeaks(xset3)

# rapport final avec statistiques de différences entre les deux classes
reporttab <- diffreport(xset5, filebase = paste(pathResult, "/Rapport_", expe, "_", polar, sep=""), mzdec=4, eicmax=5000, metlin = metlin, classeic=levels(xset5@phenoData))

# écriture du fichier Excel
dir.create(paste(pathResult, "/Rapport_", expe, "_", polar, "_diffreport/", sep=""), showWarnings = FALSE)
write.table(reporttab, paste(pathResult, "/Rapport_", expe, "_", polar, "_diffreport/resultat_", expe, "_", polar, ".xls", sep=""), sep="\t")

library(CAMERA)
# annotation version rapide?
an <- annotate(xset, platform="diagnostic", calcIso=TRUE, calcCaS=FALSE, maxcharge=3, maxiso=4, minfrac=0.5, ppm=15, mzabs=0.015, quick=FALSE, psg_list=NULL, rules=NULL, polarity=polarity)
diffreport1 <- getPeaklist(an)

# diffreport <- annotateDiffreport(xsg, pval_th=0.05, fc=0.1, nSlaves=8, calcIso=TRUE, calcCaS=FALSE, maxcharge=3, maxiso=4, minfrac=0.5, ppm=15, mzabs=0.015, quick=FALSE, psg_list=NULL, rules=NULL, polarity=polarity, sortpval=FALSE)
diffreport <- cbind(reporttab, diffreport1[, c("isotopes", "adduct", "pcgroup")])
write.table(diffreport, file=paste(pathResult, "/result_", expe, "_", polar, "_CAMERA_diffreport-fast.xls", sep=""), row.names=FALSE, sep="\t")

library(FactoMineR)
pca3 <- PCA(t(matacp), axes=c(1,2))
pca3 <- PCA(t(matacp), axes=c(1,3))
pca3 <- PCA(t(matacp), axes=c(2,3))
pca4 <- PCA(t(matacplog2))

# -- output png --
# Percentage of variance
png("percentage_of_variance.png", width =800, height = 400);
barplot(resPCA$eig$per, xlab="Components", ylab="percentage of variance");
dev.off()

png("eigenvalue.png", width =800, height = 400);
barplot(resPCA$eig$eig, xlab="Components", ylab="eigenvalue");
dev.off()

library(ctc)
# -- Normalization: logratio --
if (normalization) {
  data=t(scale(t(data)))
}
```

● Command line tools

+ represent almost the majority of scientific tools

+ good parameters completeness

+ can be executed on high performance computers

+ automatable, workflowsable, ...

- minimum linux knowledge is required

- crucial lack of ergonomics

```
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-ro
#$ -m bea
#$ -V
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err
```



```
tophat2 panda_v1
-GTF ../input/pa
-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
```

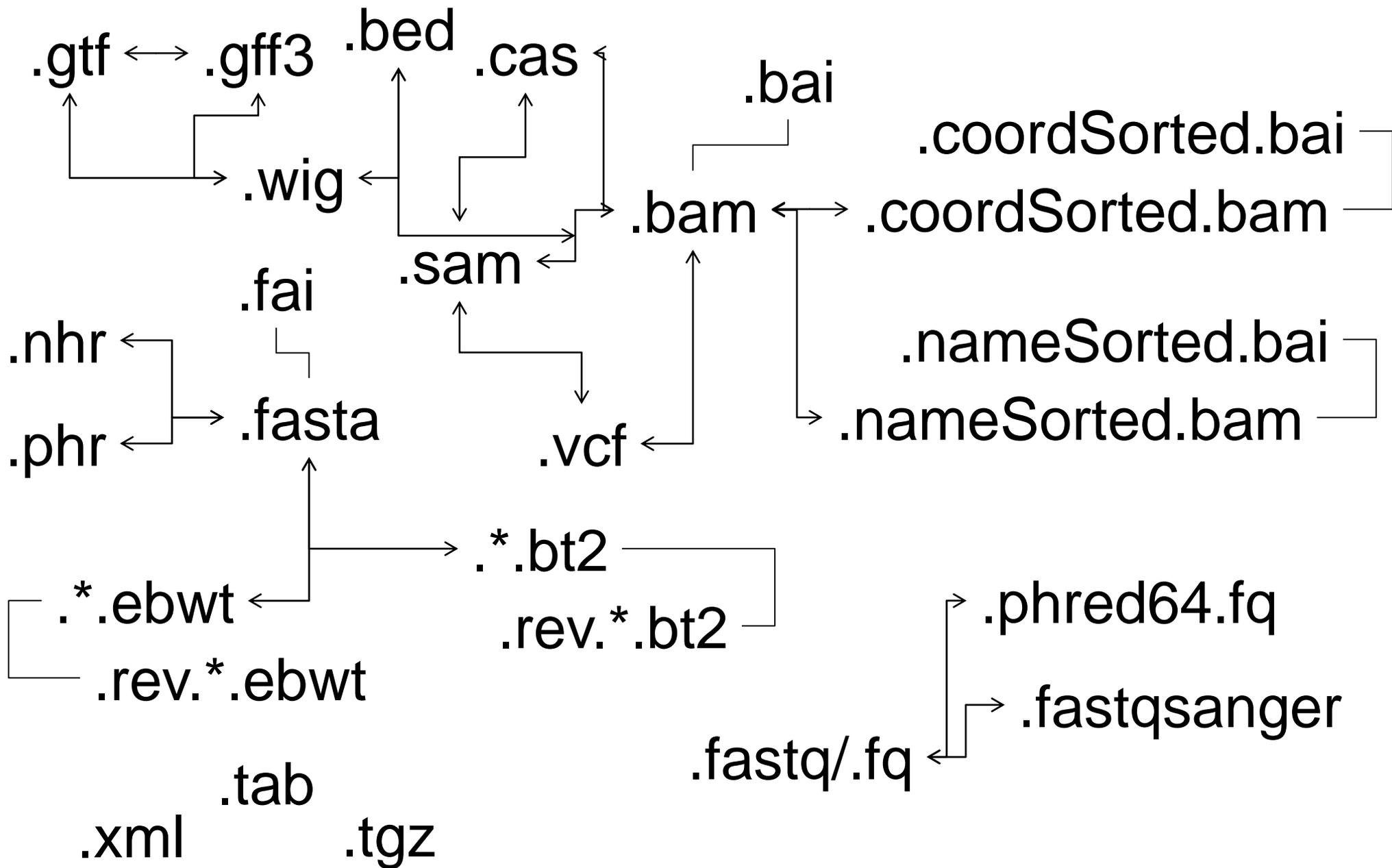
.qsub

l.bam

NOOOOOOOO!

prep_reads.info tmp

Introduction / problematic





Select your level:

Level I



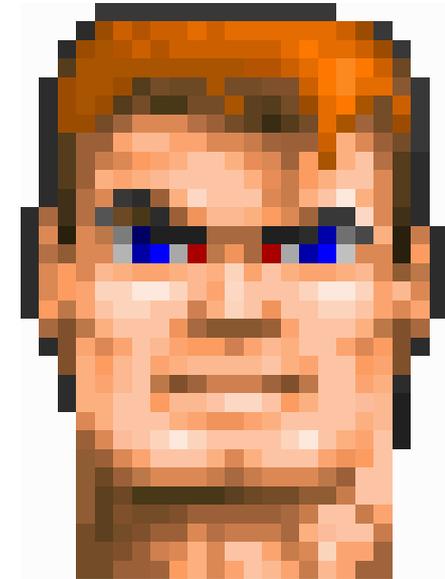
« I want to know the gene expression »

Level 2



« I want to map my reads on a reference genome and count them »

Level 3



« I want to launch the tools tophat2 and cufflinks.
I have fastq files and my genome in fasta and gtf. »

Level 4



« I want 1TB for my project. I will launch tophat2 through SSH on the cluster in multi-thread mode.

Next I want to submit the bam file to my genome with cufflinks.

Except that, I will manage :P”

Level 5



« I have a bunch of cool tools!
But I'm the only one who can launch them.

Comments? »

Galaxy

- Web-based platform for computational biomedical research (analysis and data integration)
 - Developed at Penn State, Johns Hopkins, OHSU and Cleveland Clinic with substantial outside contributions
 - Open source under Academic Free License
- More than 6,500 citations
- More than 125 [public Galaxy resources](#)
 - 100+ public servers, many more non-public
 - Both general-purpose and domain-specific



INTRODUCTION / GALAXY

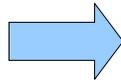
Why Galaxy ?

- Accessibility
- Reproductibility
- Transparency

- Galaxy it's ...
 - A web-based interface
 - No need to execute a command line through a terminal
 - Programming or scripting skills are not required
 - Submission of jobs is transparent through a high performance computer cluster
 - Secure histories and data manager
 - A sharing system for data and protocols
 - Tool-boxes of several bioinformatics fields
 - NGS
 - Metabolomics
 - Statistics
 - Chemistry
 - Image analysis
 - Etc ...



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 / Galaxy

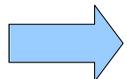
```
[login@n0 ~]$ cdprojet
[login@n0 login]$ cd 13-07-29-panda/tmp/mapping
[login@n0 mapping]$ cat tophat.qsub
#!/bin/bash
#$ -S /bin/bash
#$ -M login@sb-roscoff.fr
#$ -m bea
#$ -v
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err

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



Galaxy
PROJECT

```
[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 interface

Galaxy / 4 / Metabolomics

Analyze Data Workflow Shared Data Visualization Admin Help User

Using -993344424 b

Tools

search tools

Upload File from your computer

Export Data

LC-MS

Format Conversion

Preprocessing

Normalisation

Batch_correction Corrects intensities for signal drift and batch-effects

Determine_batch_correction to choose between linear, lowess and loess methods

Transformation Transforms the dataMatrix intensity values

Quality Control

Statistical Analysis

Annotation

GC-MS

Preprocessing

Normalisation

Quality Control

Statistical Analysis

Annotation

NMR

Preprocessing

Normalisation

Batch_correction (version 2.0.0)

Data Matrix file :

Sample metadata file :
must contain at least the three following columns: 'batch' + 'injectionOrder' + 'sampleType'

Variable metadata file :

Type of regression model :
To select between linear or non-linear (lowess or loess) methods to be used in Van der Kloet algorithm ; when using loess, you can choose to use pools or samples to model batch effect.

Factor of interest :
column name of factor of interest (often a biological factor); if none, leave 'batch'

Level of details for plots :
Amount of plots in the pdf file output. See Help section for more details.

Execute

1 Authors
Jean-Francois Martin - PF MetaToul-AXIOM ; INRA ; MetaboHUB (for original version of this tool and overall development of the R script)

1 Contributors
Melanie Petera - PFEM ; INRA ; MetaboHUB (for R wrapper and R script improvement)
Etienne Thevenot - LIST/LADIS ; CEA ; MetaboHUB (for R script and wrapper concerning "all loess pool" and "all loess sample" methods)

History

search datasets

Sacuri Zip
19 shown
289.7 MB

19:
[xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv \(Xdiffreport\)](#)

18:
[xset.group.retcor.group.fillPeaks.annotate.negative.Rdata](#)

17:
[xset.group.retcor.group.fillPeaks.annotate.dataMatrix.tsv](#)

16:
[xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv](#)

15:
[xset.group.retcor.group.fillPeaks.RData](#)

14:
[xset.group.retcor.group.Rplots.pdf](#)

13:
[xset.group.retcor.group.RData](#)

12:
[xset.group.retcor.BPCs_corrected.pdf](#)

11:

Galaxy interface

Top menu

The screenshot displays the Galaxy web interface. At the top, the navigation bar shows 'Galaxy / 4 / Metabolomics' and various menu items: 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. The user is logged in as 'Using -993344424 b'. On the left, a 'Tools' sidebar lists categories like 'LC-MS', 'Format Conversion', 'Preprocessing', and 'Normalisation', with 'Batch_correction' selected. The main workspace shows the configuration for the 'Batch_correction (version 2.0.0)' tool. The configuration includes fields for 'Data Matrix file', 'Sample metadata file', 'Variable metadata file', 'Type of regression model' (set to 'linear'), 'Factor of interest' (set to 'batch'), and 'Level of details for plots' (set to 'basic'). An 'Execute' button is visible at the bottom of the configuration area. Below the configuration, the 'Authors' and 'Contributors' sections are displayed. On the right, a 'History' panel shows a list of datasets, including 'Sacuri Zip' and several files generated by the 'Batch_correction' tool, such as 'xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv (Xdiffreport)' and 'xset.group.retcor.group.fillPeaks.annotate.negative.Rdata'.

Galaxy / 4 / Metabolomics Analyze Data Workflow Shared Data Visualization Admin Help User Using -993344424 b

Tools search tools

Batch_correction (version 2.0.0)

Data Matrix file : 17: xset.group.retcor.group.fillPeaks.annotate.dataMatrix.tsv

Sample metadata file : 3: sampleMetadata.tsv
must contain at least the three following columns: 'batch' + 'injectionOrder' + 'sampleType'

Variable metadata file : 16: xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv

Type of regression model : linear
To select between linear or non-linear (lowess or loess) methods to be used in Van der Kloet algorithm ; when using loess, you can choose to use pools or samples to model batch effect.

Factor of interest : batch
column name of factor of interest (often a biological factor); if none, leave 'batch'

Level of details for plots : basic
Amount of plots in the pdf file output. See Help section for more details.

Execute

1 Authors
Jean-Francois Martin - PF MetaToul-AXIOM ; INRA ; MetaboHUB (for original version of this tool and overall development of the R script)

1 Contributors
Melanie Petera - PFEM ; INRA ; MetaboHUB (for R wrapper and R script improvement)
Etienne Thevenot - LIST/LADIS ; CEA ; MetaboHUB (for R script and wrapper concerning "all loess pool" and "all loess sample" methods)

History search datasets

Sacuri Zip
19 shown
289.7 MB

19: xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv (Xdiffreport)

18: xset.group.retcor.group.fillPeaks.annotate.negative.Rdata

17: xset.group.retcor.group.fillPeaks.annotate.dataMatrix.tsv

16: xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv

15: xset.group.retcor.group.fillPeaks.RData

14: xset.group.retcor.group.Rplots.pdf

13: xset.group.retcor.group.RData

12: xset.group.retcor.BPCs_corrected.pdf

11:

Galaxy interface

Tool list

Galaxy / 4 / Metabolomics Analyze Data Workflow Shared Data Visualization Admin Help User Using -993344424 b

Tools search tools

Upload File from your computer
Export Data

LC-MS
Format Conversion
Preprocessing
Normalisation

Batch_correction Corrects intensities for signal drift and batch-effects

Determine_batch_correction to choose between linear, lowess and loess methods

Transformation Transforms the dataMatrix intensity values

Quality Control
Statistical Analysis
Annotation

GC-MS
Preprocessing
Normalisation
Quality Control
Statistical Analysis
Annotation

NMR
Preprocessing
Normalisation

Batch_correction (version 2.0.0)

Data Matrix file : 17: xset.group.retcor.group.fillPeaks.annotate.dataMatrix.tsv

Sample metadata file : 3: sampleMetadata.tsv
must contain at least the three following columns: 'batch' + 'injectionOrder' + 'sampleType'

Variable metadata file : 16: xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv

Type of regression model : linear
To select between linear or non-linear (lowess or loess) methods to be used in Van der Kloet algorithm ; when using loess, you can choose to use pools or samples to model batch effect.

Factor of interest : batch
column name of factor of interest (often a biological factor); if none, leave 'batch'

Level of details for plots : basic
Amount of plots in the pdf file output. See Help section for more details.

Execute

1 Authors
Jean-Francois Martin - PF MetaToul-AXIOM ; INRA ; MetaboHUB (for original version of this tool and overall development of the R script)

1 Contributors
Melanie Petera - PFEM ; INRA ; MetaboHUB (for R wrapper and R script improvement)
Etienne Thevenot - LIST/LADIS ; CEA ; MetaboHUB (for R script and wrapper concerning "all loess pool" and "all loess sample" methods)

History search datasets

Sacuri Zip
19 shown
289.7 MB

19: xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv (Xdiffreport)

18: xset.group.retcor.group.fillPeaks.annotate.negative.Rdata

17: xset.group.retcor.group.fillPeaks.annotate.dataMatrix.tsv

16: xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv

15: xset.group.retcor.group.fillPeaks.RData

14: xset.group.retcor.group.Rplots.pdf

13: xset.group.retcor.group.RData

12: xset.group.retcor.BPCs_corrected.pdf

11:

Galaxy interface

Web forms / dataset visualization / miscellaneous information

Galaxy / 4 / Metabolomics Analyze Data Workflow Shared Data Visualization Admin Help User Using -993344424 b

Tools search tools

Upload File from your computer
Export Data

LC-MS
Format Conversion
Preprocessing
Normalisation

Batch_correction Corrects intensities for signal drift and batch-effects
Determine_batch_correction to choose between linear, lowess and loess methods
Transformation Transforms the dataMatrix intensity values

Quality Control
Statistical Analysis
Annotation

GC-MS
Preprocessing
Normalisation
Quality Control
Statistical Analysis
Annotation

NMR
Preprocessing
Normalisation

Batch_correction (version 2.0.0)

Data Matrix file : 17: xset.group.retcor.group.fillPeaks.annotate.dataMatrix.tsv

Sample metadata file : 3: sampleMetadata.tsv
must contain at least the three following columns: 'batch' + 'injectionOrder' + 'sampleType'

Variable metadata file : 16: xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv

Type of regression model : linear
To select between linear or non-linear (lowess or loess) methods to be used in Van der Kloet algorithm ; when using loess, you can choose to use pools or samples to model batch effect.

Factor of interest : batch
column name of factor of interest (often a biological factor); if none, leave 'batch'

Level of details for plots : basic
Amount of plots in the pdf file output. See Help section for more details.

Execute

1 Authors
Jean-Francois Martin - PF MetaToul-AXIOM ; INRA ; MetaboHUB (for original version of this tool and overall development of the R script)

1 Contributors
Melanie Petera - PFEM ; INRA ; MetaboHUB (for R wrapper and R script improvement)
Etienne Thevenot - LIST/LADIS ; CEA ; MetaboHUB (for R script and wrapper concerning "all loess pool" and "all loess sample" methods)

History search datasets

Sacuri Zip
19 shown
289.7 MB

19: xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv (Xdiffreport)

18: xset.group.retcor.group.fillPeaks.annotate.negative.Rdata

17: xset.group.retcor.group.fillPeaks.annotate.dataMatrix.tsv

16: xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv

15: xset.group.retcor.group.fillPeaks.RData

14: xset.group.retcor.group.Rplots.pdf

13: xset.group.retcor.group.RData

12: xset.group.retcor.BPCs_corrected.pdf

11:

Galaxy interface

History

Galaxy / 4 / Metabolomics

Analyze Data Workflow Shared Data Visualization Admin Help User

Using -993344424 b

Tools

search tools

Upload File from your computer

Export Data

LC-MS

Format Conversion

Preprocessing

Normalisation

Batch_correction Corrects intensities for signal drift and batch-effects

Determine_batch_correction to choose between linear, lowess and loess methods

Transformation Transforms the dataMatrix intensity values

Quality Control

Statistical Analysis

Annotation

GC-MS

Preprocessing

Normalisation

Quality Control

Statistical Analysis

Annotation

NMR

Preprocessing

Normalisation

Batch_correction (version 2.0.0)

Data Matrix file :

Sample metadata file :
must contain at least the three following columns: 'batch' + 'injectionOrder' + 'sampleType'

Variable metadata file :

Type of regression model :
To select between linear or non-linear (lowess or loess) methods to be used in Van der Kloet algorithm ; when using loess, you can choose to use pools or samples to model batch effect.

Factor of interest :
column name of factor of interest (often a biological factor); if none, leave 'batch'

Level of details for plots :
Amount of plots in the pdf file output. See Help section for more details.

Execute

1 Authors
Jean-Francois Martin - PF MetaToul-AXIOM ; INRA ; MetaboHUB (for original version of this tool and overall development of the R script)

1 Contributors
Melanie Petera - PFEM ; INRA ; MetaboHUB (for R wrapper and R script improvement)
Etienne Thevenot - LIST/LADIS ; CEA ; MetaboHUB (for R script and wrapper concerning "all loess pool" and "all loess sample" methods)

History

search datasets

Sacuri Zip
19 shown
289.7 MB

19:
[xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv \(Xdiffreport\)](#)

18:
[xset.group.retcor.group.fillPeaks.annotate.negative.Rdata](#)

17:
[xset.group.retcor.group.fillPeaks.annotate.dataMatrix.tsv](#)

16:
[xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv](#)

15:
[xset.group.retcor.group.fillPeaks.RData](#)

14:
[xset.group.retcor.group.Rplots.pdf](#)

13:
[xset.group.retcor.group.RData](#)

12:
[xset.group.retcor.BPCs_corrected.pdf](#)

11:

GET HELP

Galaxy / ABiMS

Analyze Data Workflow Shared Data Visualization Admin Help User
Using 0%

Tools

Get Data

Send Data

Collection Operations

COMMON TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

Extract Features

Fetch Sequences

Statistics

Graph/Display Data

Fasta Fastq Manipulation

COMMON NGS TOOLS

NGS:Samtools

NGS:Mapping

NGS:Bedtools

NGS:Picard Tools

SEARCHING TOOLS

Diamond

✓ Welcome to galaxy3.sb-roscoff.fr

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



Analyses and Bioinformatics for Marine Science

▸ Changelog

▸ Tutorials

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

eba 2016 sartools
42 shown

1.59 MB

- 62: SARTools DESeq2 R objects (.RData)**
- 61: SARTools DESeq2 R log**
- 60: SARTools DESeq2 figures**
- 59: SARTools DESeq2 tables**
- 58: SARTools DESeq2 report**
- 57: SARTools edgeR R objects (.RData)**
- 56: SARTools edgeR R log**
- 55: SARTools edgeR figures**

Galaxy / ABiMS

 Analyze Data Workflow Shared Data ▾ Visualization ▾ Admin Help ▾ User ▾ Using 0%

Tools

Get Data

Send Data

Collection Operations

COMMON TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

Extract Features

Fetch Sequences

Statistics

Graph/Display Data

Fasta Fastq Manipulation

COMMON NGS TOOLS

NGS:Samtools

NGS:Mapping

NGS:Bedtools

NGS:Picard Tools

SEARCHING TOOLS

Diamond

✓ **Welcome to galaxy3.sb-roscoff.fr**

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



Analyses and Bioinformatics for Marine Science

▸ Changelog

▸ Tutorials

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

eba 2016 sartools
42 shown

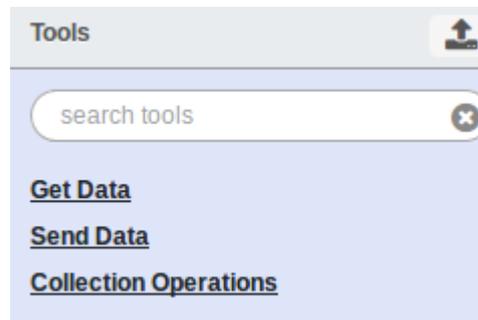
1.59 MB

- 62: SARTools DESeq2 R objects (.RData)**
- 61: SARTools DESeq2 R log**
- 60: SARTools DESeq2 figures**
- 59: SARTools DESeq2 tables**
- 58: SARTools DESeq2 report**
- 57: SARTools edgeR R objects (.RData)**
- 56: SARTools edgeR R log**
- 55: SARTools edgeR figures**

DATA IMPORT

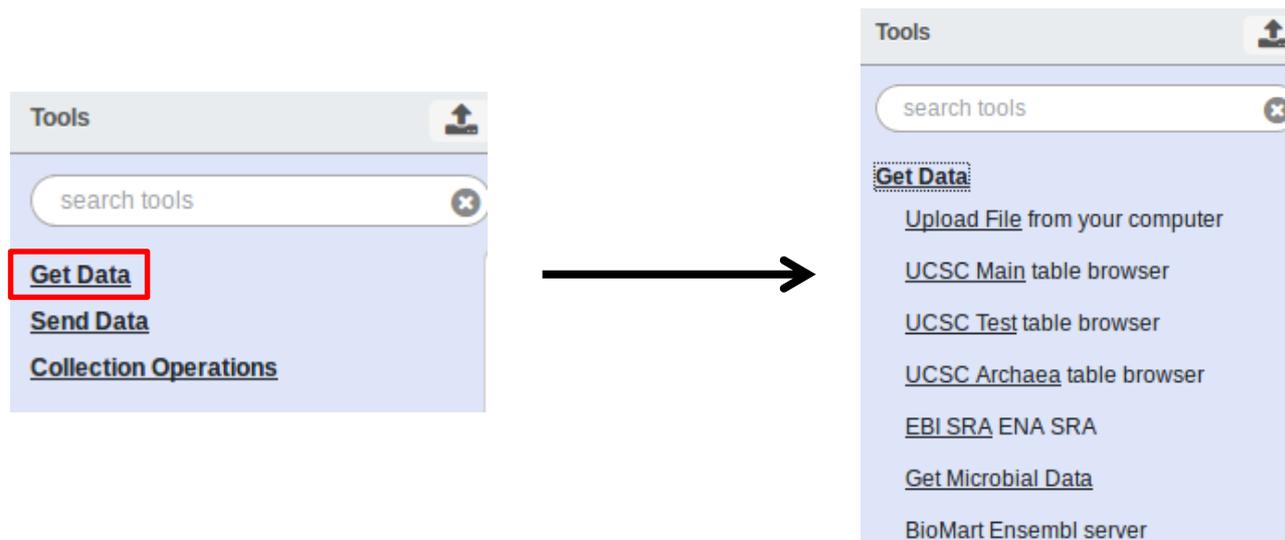
- Copy/paste from a file
- Upload data from your local computer
- Upload data from internet using URL
- Upload data from FTP
- Upload data from online databases: EBI ENA, UCSC, BioMart, etc.
- Import from Shared Data (libraries, histories, workflows)

- Copy/paste from a file
- Upload data from your local computer
- Upload data from internet using URL
- Upload data from FTP
- Upload data from online databases: EBI ENA, UCSC, BioMart, etc.
- Import from Shared Data (libraries, histories, workflows)



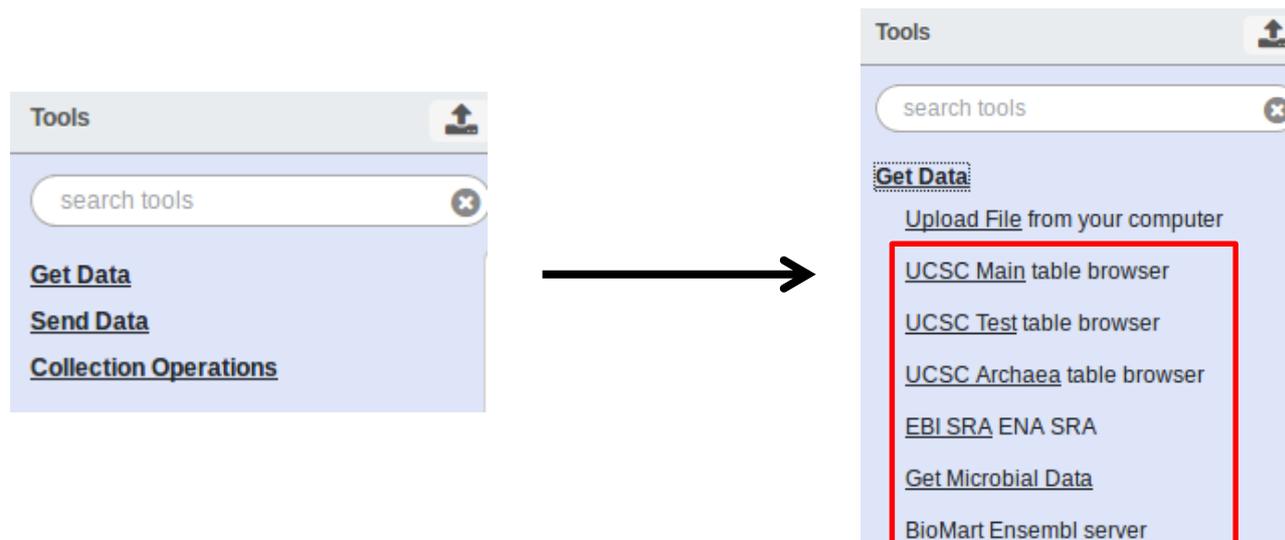
Data import

- Copy/paste from a file
- Upload data from your local computer
- Upload data from internet using URL
- Upload data from FTP
- Upload data from online databases: EBI ENA, UCSC, BioMart, etc.
- Import from Shared Data (libraries, histories, workflows)



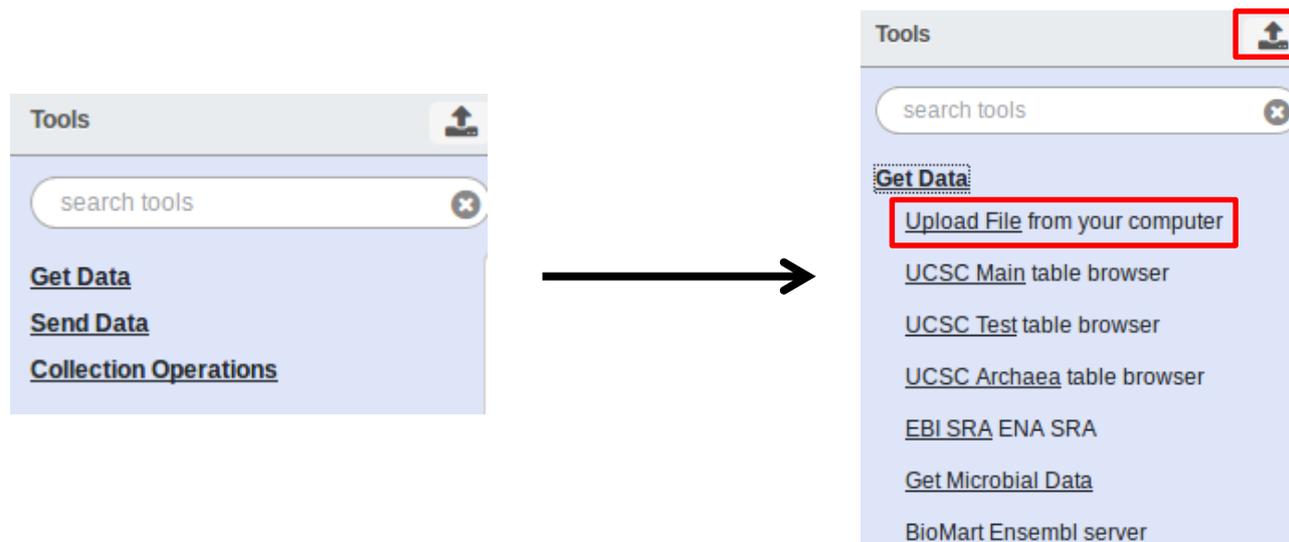
Data import

- Copy/paste from a file
- Upload data from your local computer
- Upload data from internet using URL
- Upload data from FTP
- Upload data from online databases: EBI ENA, UCSC, BioMart, etc.
- Import from Shared Data (libraries, histories, workflows)



Data import

- Copy/paste from a file
- Upload data from your local computer
- Upload data from internet using URL
- Upload data from FTP
- Upload data from online databases: EBI ENA, UCSC, BioMart, etc.
- Import from Shared Data (libraries, histories, workflows)



Data import

The screenshot shows the Galaxy 4 / Metabolomics website interface. The browser address bar displays 'galaxy.workflow4metabolomics.org'. The main navigation bar includes 'Galaxy / 4 / Metabolomics', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. A 'Tools' sidebar on the left lists various analysis categories such as 'Upload File from your computer', 'Export Data', 'LC-MS', 'Preprocessing', 'Normalisation', 'Statistical Analysis', 'Annotation', 'GC-MS', 'NMR', and 'COMMON TOOLS'. A tooltip over the 'Tools' icon reads 'Download from URL or upload files from disk'. The main content area features a green banner for 'Welcome to workflow4metabolomics.org v2.0' with a publication citation: 'Publication: Franck Giacomoni, Gildas Le Corguillé, Mishari Monsoor, Marion Landi, Pierre Pericard, Mélanie Pétéra, Christophe Duperier, Marie Tremblay-Franco, Jean-François Martin, Daniel Jacob, Sophie Goulitquer, Etienne A. Thévenot and Christophe Caron (2014). Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics. Bioinformatics doi:10.1093/bioinformatics/btu813'. Below this is a 'Latest news' section with three entries dated 01/06/2015, 21/09/2015, and 19/12/2014. A 'History' panel on the right shows 'Unnamed history' with '0 bytes' and a message: 'This history is empty. You can load your own data or get data from an external source'. The bottom of the page features a workflow diagram with sections for 'LC/MS', 'MS', and 'Common'. The Windows taskbar at the bottom shows various application icons and the system clock indicating 10:23 AM on 6/5/2015.

Data import

The screenshot shows a web browser window with the URL `galaxy.workflow4metabolomics.org`. The page title is "Galaxy / 4 / Metabolomics". The browser's address bar and navigation icons are visible at the top. The main content area features a navigation menu on the left with categories like "Tools", "LC-MS", "GC-MS", and "NMR", each with sub-links for "Preprocessing", "Normalisation", "Quality Control", and "Statistical Analysis". A central dialog box is open, titled "Download data directly from web or upload files from your disk". It contains a large dashed-line box for file upload, with the text "You can Drag & Drop files into this box." below it. At the bottom of the dialog are buttons for "Choose local file", "Choose FTP file", "Paste/Fetch data", "Start", "Pause", "Reset", and "Close". The background shows a workflow diagram with various processing steps. The Windows taskbar at the bottom includes icons for Internet Explorer, File Explorer, Word, Excel, PowerPoint, and other applications. The system tray shows the date and time as "10:23 AM 6/5/2015".

Data import

Copy / Paste data

Galaxy / 4 / Metabolomics

galaxy.workflow4metabolomics.org

Galaxy / 4 / Metabolomics Analyze Data Workflow Shared Data Visualization Admin Help User

Tools

search tools

Upload File from your computer

Export Data

LC-MS

Preprocessing

Normalisation

Statistical Analysis

Annotation

GC-MS

Preprocessing

Normalisation

Quality Control

Statistical Analysis

Annotation

NMR

Preprocessing

Normalisation

Quality Control

Statistical Analysis

COMMON TOOLS

Text Manipulation

Filter and Sort

Download data directly from web or upload files from your disk

Name	Size	Type	Genome	Settings	Status
New File	0.9 KB	Auto-detect	unspecified (?)		

You can tell Galaxy to download data from web by entering URL in this box (one per line)

sampleMetadata class polarity batch
Blanc15 blank negative 1
Blanc09 blank negative 1

Upload configuration

- Convert spaces to tabs
- Use POSIX standard

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

Choose local file Choose FTP file Paste/Fetch data Start Pause Reset Close

10:23 AM 6/5/2015

Data import

From local files

The screenshot shows a web browser window with the URL `galaxy.workflow4metabolomics.org`. The page title is "Galaxy / 4 / Metabolomics". A modal dialog box is open in the center, titled "Download data directly from web or upload files from your disk". The dialog contains a large dashed rectangular area for file upload, with the text "You can Drag & Drop files into this box." below it. At the bottom of the dialog, there are several buttons: "Choose local file" (highlighted with a red box), "Choose FTP file", "Paste/Fetch data", "Start", "Pause", "Reset", and "Close". The background shows the Galaxy interface with a sidebar of tools categorized by LC-MS, GC-MS, and NMR, and a main workspace area.

ARMS

Data import

From local files

The screenshot shows the Galaxy web interface in a browser window. The address bar contains the URL `galaxy.workflow4metabolomics.org`. The page title is `Galaxy / 4 / Metabolomics`. The main content area is a dialog box titled `Download data directly from web or upload files from your disk`. The dialog has a dashed border and contains a folder icon with a `Move` button. Below the dialog is a text prompt: `You can Drag & Drop files into this box.` At the bottom of the dialog are several buttons: `Choose local file`, `Choose FTP file`, `Paste/Fetch data`, `Start`, `Pause`, `Reset`, and `Close`. The background shows the Galaxy interface with a sidebar of tool categories and a workflow editor.

ARMS

Data import

From local files

Galaxy / 4 / Metabolomics

galaxy.workflow4metabolomics.org

Galaxy / 4 / Metabolomics

Analyze Data Workflow Shared Data Visualization Admin Help User

Tools

search tools

Upload File from your computer

Export Data

LC-MS

Preprocessing

Normalisation

Statistical Analysis

Annotation

GC-MS

Preprocessing

Normalisation

Quality Control

Statistical Analysis

Annotation

NMR

Preprocessing

Normalisation

Quality Control

Statistical Analysis

COMMON TOOLS

Text Manipulation

Filter and Sort

Download data directly from web or upload files from your disk

Name	Size	Type	Genome	Settings	Status
 sacuri.zip	0.2 GB	Auto-det...	unspecified (?)		

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

Choose local file Choose FTP file Paste/Fetch data **Start** Pause Reset Close

Using 2.5 GB

history

search datasets

unnamed history

bytes

This history is empty. You can [load your own data](#) or [get data from an external source](#)

ARMS

FR 10:23 AM 6/5/2015

Data import

From local files

Galaxy / 4 / Metabolomics

galaxy.workflow4metabolomics.org

Galaxy / 4 / Metabolomics

Analyze Data Workflow Shared Data Visualization Admin Help User

Tools

search tools

Upload File from your computer

Export Data

LC-MS

Preprocessing

Normalisation

Statistical Analysis

Annotation

GC-MS

Preprocessing

Normalisation

Quality Control

Statistical Analysis

Annotation

NMR

Preprocessing

Normalisation

Quality Control

Statistical Analysis

COMMON TOOLS

Text Manipulation

Filter and Sort

Download data directly from web or upload files from your disk

Name	Size	Type	Genome	Settings	Status
sacuri.zip	0.2 GB	Auto-det...	unspecified (?)		

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

Choose local file Choose FTP file Paste/Fetch data Start Pause Reset Close

Using 2.5 GB

history

search datasets

unnamed history

bytes

This history is empty. You can [load your own data](#) or [get data from an external source](#)

ARMS

FR 10:23 AM 6/5/2015

Data import

From local files

Galaxy / 4 / Metabolomics

galaxy.workflow4metabolomics.org

Galaxy / 4 / Metabolomics

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 2.5 GB

Tools

search tools

Upload File from your computer

Export Data

LC-MS

Preprocessing

Normalisation

Statistical Analysis

Annotation

GC-MS

Preprocessing

Normalisation

Quality Control

Statistical Analysis

Annotation

NMR

Preprocessing

Normalisation

Quality Control

Statistical Analysis

COMMON TOOLS

Text Manipulation

Filter and Sort

Download data directly from web or upload files from your disk

Name	Size	Type	Genome	Settings	Status
sacuri.zip	0.2 GB	Auto-det...	unspecified (?)	⚙️	50%

Please wait...1 out of 1 remaining.

Choose local file Choose FTP file Paste/Fetch data Start Pause Reset Close

ARMS

FR 10:23 AM 6/5/2015

Data import

From local files

The screenshot shows the Galaxy 4 Metabolomics web interface. A dialog box titled "Download data directly from web or upload files from your disk" is open. Inside the dialog, a table lists the file "sacuri.zip" with a size of 0.2 GB, type "Auto-det...", genome "unspecified (?)", and a status of 100% with a checkmark. Below the table, a dashed box contains the text "You can Drag & Drop files into this box." At the bottom of the dialog are buttons for "Choose local file", "Choose FTP file", "Paste/Fetch data", "Start", "Pause", "Reset", and "Close". The background shows the Galaxy interface with a sidebar of tool categories and a main workspace area.

Name	Size	Type	Genome	Settings	Status
sacuri.zip	0.2 GB	Auto-det...	unspecified (?)	⚙️	100% ✓

You can Drag & Drop files into this box.

Choose local file Choose FTP file Paste/Fetch data Start Pause Reset Close



Data import

From local files

The screenshot displays the Galaxy workflow4metabolomics.org v2.0 web interface. The browser address bar shows the URL galaxy.workflow4metabolomics.org. The main navigation bar includes 'Galaxy / 4 / Metabolomics' and various menu items like 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. The interface is divided into several sections:

- Tools Panel (Left):** A sidebar with a search bar and categories for 'LC-MS', 'GC-MS', and 'NMR'. Each category lists sub-tools such as 'Preprocessing', 'Normalisation', 'Quality Control', and 'Statistical Analysis'.
- Welcome Message (Top Center):** A green banner with a checkmark icon. Text: 'Welcome to workflow4metabolomics.org v2.0'. Below it, a publication reference: 'Publication: Franck Giacomoni, Gildas Le Corguillé, Mishari Monsoor, Marion Landi, Pierre Pericard, Mélanie Pétéra, Christophe Duperier, Marie Tremblay-Franco, Jean-François Martin, Daniel Jacob, Sophie Goulitquer, Etienne A. Thévenot and Christophe Caron (2014). Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics. Bioinformatics doi:10.1093/bioinformatics/btu813'. A help and support email is provided: support@workflow4metabolomics.org.
- Latest News (Middle):** A light blue box with an information icon. Title: 'Latest news'. News items include: '01/06/2015 - Workflow4Metabolomics v2.0 starts today - Check the changelog section below', '01/06/2015 - The W4M 2.0 release is presented in the June 2015 MetaboNews Spotlight [link]', '21/09/2015 - Ecole-chercheurs : Traitement des données métabolomiques sur l'infrastructure online Workflow4Metabolomics (21-25 Sept. 2015) [in French] / Roscoff, France', and '19/12/2014 - W4M publication in Bioinformatics is now available - Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics'. Below the news are links for 'Changelog', 'Tutorials', and 'Past events'.
- History Panel (Right):** A sidebar titled 'History' with a search bar. It shows 'Unnamed history' with '0 bytes'. Below that, a dataset entry '1: sacuri.zip' is highlighted with a mouse cursor. The entry includes icons for viewing, editing, and deleting.
- Workflow Diagrams (Bottom):** A section showing workflow diagrams for 'LC/MS', 'MS', and 'Common' categories.

The Windows taskbar at the bottom shows various application icons, including Internet Explorer, Word, Excel, PowerPoint, and a clock displaying 10:23 AM on 6/5/2015.

Data import

From local files

The screenshot displays the Galaxy 4 Metabolomics web interface. The browser address bar shows `galaxy.workflow4metabolomics.org`. The main navigation bar includes "Galaxy / 4 / Metabolomics" and various menu items like "Analyze Data", "Workflow", "Shared Data", "Visualization", "Admin", "Help", and "User".

On the left sidebar, under "Tools", there is a search bar and a section for "Upload File from your computer". Below this, there are categories for LC-MS, GC-MS, and NMR, each with sub-sections for "Preprocessing", "Normalisation", "Quality Control", "Statistical Analysis", and "Annotation". At the bottom of the sidebar, there are "COMMON TOOLS" including "Text Manipulation" and "Filter and Sort".

The main content area features a green banner with a checkmark icon and the text: "Welcome to workflow4metabolomics.org v2.0". Below this is a "Publication" section with the text: "Publication: Franck Giacomoni, Gildas Le Corguillé, Mishari Monsoor, Marion Landi, Pierre Pericard, Mélanie Pétéra, Christophe Duperier, Marie Tremblay-Franco, Jean-François Martin, Daniel Jacob, Sophie Goulitquer, Etienne A. Thévenot and Christophe Caron (2014). Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics. Bioinformatics doi:10.1093/bioinformatics/btu813". A "Help and support" link is provided below.

Below the welcome message is a "Latest news" section with an information icon. It contains three news items: "01/06/2015 - Workflow4Metabolomics v2.0 starts today - Check the changelog section below", "01/09/2015 - The W4M 2.0 release is presented in the June 2015 MetaboNews Spotlight [link]", and "21/09/2015 - Ecole-chercheurs : Traitement des données métabolomiques sur l'infrastructure online Workflow4Metabolomics (21-25 Sept. 2015) [in French] / Roscoff, France". A fourth item is dated "19/12/2014 - W4M publication in Bioinformatics is now available - Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics".

At the bottom of the main content area, there are three panels: "LC/MS", "MS", and "Common", each showing a workflow diagram with various tool icons.

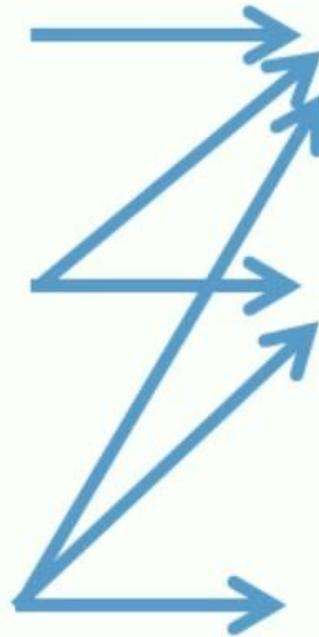
On the right sidebar, the "History" section shows a search bar and a list of datasets. The first entry is "1: sacuri.zip" with a status of "This job is currently running".

The Windows taskbar at the bottom shows several application icons, including Internet Explorer, Word, Excel, PowerPoint, and a yellow duck icon. The system tray on the right shows the date and time: "10:23 AM 6/5/2015".

Step 1: Choose a FTP Client

DATA IMPORT USING FTP

STEP 1: CHOOSE A FTP CLIENT



FileZilla

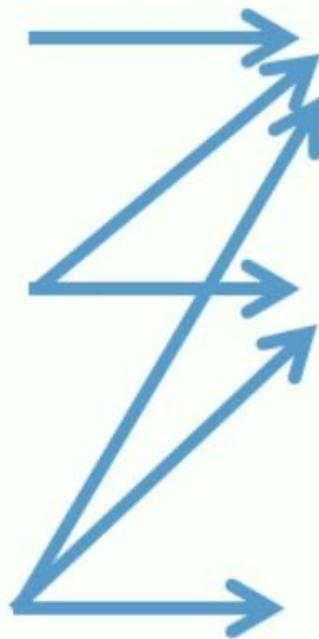


Cyberduck



WinSCP

STEP 1: CHOOSE A FTP CLIENT



Avoid:
Malwares inside

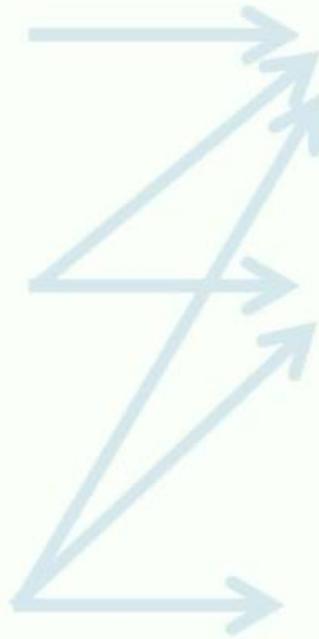


Cyberduck



WinSCP

STEP 1: CHOOSE A FTP CLIENT



FileZilla



Cyberduck



WinSCP

Step 2: Easy!

DATA IMPORT USING FTP

Data import using FTP

The screenshot shows the Galaxy / ABiMS web interface in a browser window. The address bar displays 'galaxy.sb-roscoff.fr'. The main navigation bar includes 'Galaxy / ABiMS', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. A 'Using 46%' indicator is visible in the top right corner.

The left sidebar contains a 'Tools' section with a search bar and a 'Download from URL or upload files from disk' button. Below this, there are categories of tools: 'Get Data', 'COMMON TOOLS', 'Convert Formats', 'FASTA manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Text Manipulation', 'Graphics', 'Statistics', 'EMBOSS 5 Suite', 'SEARCHING TOOLS', 'NCBI BLAST+', 'Diamond', 'Primer/Microsatellite', 'NGS TOOLS', 'NGS: BedTools', 'NGS: Mapping', 'NGS: Picard', and 'NGS: QC and manipulation'.

The main content area features a green banner with the text 'Welcome to galaxy.sb-roscoff.fr'. Below this is an 'Information' box with an 'i' icon and the text: 'For any question or request for tools or account, send an email at support.abims@sb-roscoff.fr'. The central logo for 'ABiMS' is displayed, with 'Analyses and Bioinformatics for Marine Science' underneath. To the right of the logo is the logo for 'CNRS UPRC Station Biologique Roscoff'. Below the logo are two expandable sections: 'Changelog' and 'Tutorials'. At the bottom of the main content area, there is a paragraph of text: '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.'

The right sidebar shows a 'History' section with a search bar for 'search datasets' and an 'Unnamed history' section with '0 bytes' and a message: 'This history is empty. You can load your own data or get data from an external source'.

The Windows taskbar at the bottom shows various application icons, including Internet Explorer, Chrome, File Explorer, Word, Excel, PowerPoint, and a duck icon. The system tray in the bottom right corner shows the date and time: 'FR 11:19 AM 7/31/2015'.

Data import using FTP

The screenshot shows a web browser window with the URL `galaxy.sb-roscoff.fr`. The page title is "Galaxy / ABiMS". A modal dialog box is open in the center, titled "Download data directly from web or upload files from your disk". The dialog contains a large dashed box for file upload with the text "You can Drag & Drop files into this box." Below the box are two dropdown menus: "Type (set all):" set to "Auto-detect" and "Genome (set all):" set to "unspecified (?)". At the bottom of the dialog are buttons for "Choose local file", "Choose FTP file", "Paste/Fetch data", "Start", "Pause", "Reset", and "Close". The background interface shows a sidebar with tool categories like "COMMON TOOLS", "SEARCHING TOOLS", and "NGS TOOLS". The Windows taskbar at the bottom shows the time as 11:19 AM on 7/31/2015.

Data import using FTP

The screenshot shows a web browser window with the URL `galaxy.sb-roscoff.fr`. The page title is "Galaxy / ABiMS". The navigation menu includes "Analyze Data", "Workflow", "Shared Data", "Visualization", "Admin", "Help", and "User". A modal dialog box is open with the title "Download data directly from web or upload files from your disk". Inside the dialog, there is a large dashed box with the text "You can Drag & Drop files into this box." Below this box are two dropdown menus: "Type (set all):" with "Auto-detect" selected, and "Genome (set all):" with "unspecified (?)" selected. At the bottom of the dialog are several buttons: "Choose local file", "Choose FTP file" (which is highlighted by a mouse cursor), "Paste/Fetch data", "Start", "Pause", "Reset", and "Close". The background of the browser shows a sidebar with various tool categories like "COMMON TOOLS", "SEARCHING TOOLS", and "NGS TOOLS". The Windows taskbar at the bottom shows the time as 11:19 AM on 7/31/2015.

Data import using FTP

The screenshot shows the Galaxy/ABiMS web interface. A modal dialog box is open with the title "Download data directly from web or upload files from your disk". Inside the dialog, there is a dashed box for file upload and a smaller "FTP files" sub-dialog. The "FTP files" dialog contains the text: "This Galaxy server allows you to upload files via FTP. To upload some files, log in to the FTP server at **galaxy.sb-roscoff.fr** using your Galaxy credentials (email address and password)." Below this text is a yellow warning box that says "Your FTP directory does not contain any files." At the bottom of the main dialog, there are several buttons: "Choose local file", "Choose FTP file" (which is highlighted by a mouse cursor), "Paste/Fetch data", "Start", "Pause", "Reset", and "Close". The background shows the Galaxy/ABiMS navigation menu with categories like "Tools", "Get Data", "COMMON TOOLS", "SEARCHING TOOLS", and "NGS TOOLS".

Data import using FTP

The screenshot shows a web browser window with the URL `galaxy.sb-roscoff.fr`. The page title is "Galaxy / ABiMS". The main content area is a dialog box titled "Download data directly from web or upload files from your disk". Inside this dialog, there is a dashed box for file upload with the text "You can Drag & Drop files into this box." Below this, there is a sub-dialog titled "FTP files" with the following text: "This Galaxy server allows you to upload files via FTP. To upload some files, log in to the FTP server at **galaxy.sb-roscoff.fr** using your Galaxy credentials (email address and password)." Below the text is a yellow warning box that says "Your FTP directory does not contain any files." At the bottom of the main dialog, there are several buttons: "Choose local file", "Choose FTP file", "Paste/Fetch data", "Start", "Pause", "Reset", and "Close". The background shows the Galaxy/ABiMS navigation menu with categories like "Tools", "Get Data", "COMMON TOOLS", "SEARCHING TOOLS", and "NGS TOOLS". The Windows taskbar at the bottom shows various application icons and the system clock indicating 11:19 AM on 7/31/2015.

Data import using FTP

The screenshot shows a Windows desktop environment. In the background, a web browser window displays the Galaxy/ABiMS interface at galaxy.sb-roscoff.fr. The browser's address bar shows the URL, and the page title is "Galaxy / ABiMS". The interface includes a navigation menu with options like "Analyze Data", "Workflow", "Shared Data", "Visualization", "Admin", "Help", and "User". A central panel titled "Download data directly from web or upload files from your disk" contains a large dashed box for file uploads and a message: "You can Drag & Drop files into this box." Below this, there is a text box explaining that files can be uploaded via FTP by logging into galaxy.sb-roscoff.fr. A yellow warning box states: "Your FTP directory does not contain any files." At the bottom of the panel, there are buttons for "local file", "Choose FTP file", "Paste/Fetch data", "Start", "Pause", "Reset", and "Close".

In the foreground, a Cyberduck window is open, titled "Cyberduck" and marked as "Unregistered". The window has a menu bar (File, Edit, View, Go, Bookmark, Window, Help) and a toolbar with icons for "Open Connection", "Quick Connect", "Action", "Get Info", and "Refresh". The main area of the window is currently empty. The Windows taskbar at the bottom shows various application icons, including Internet Explorer, Chrome, File Explorer, Word, Excel, PowerPoint, and a duck icon. The system tray in the bottom right corner shows the date and time as "FR 11:19 AM 7/31/2015".

Data import using FTP

The image shows a Windows desktop environment with a web browser displaying the Galaxy/ABiMS interface. The browser's address bar shows `galaxy.sb-roscoff.fr`. The web page has a navigation menu with items like "Analyze Data", "Workflow", "Shared Data", "Visualization", "Admin", "Help", and "User". A central panel is titled "Download data directly from web or upload files from your disk" and contains a large dashed box for file uploads. A message below the box says "You can Drag & Drop files into this box." and another message below that says "Your FTP directory does not contain any files." At the bottom of the web page, there are buttons for "local file", "Choose FTP file", "Paste/Fetch data", "Start", "Pause", "Reset", and "Close".

Overlaid on the browser is the Cyberduck application window. The "Open connection" dialog is active, showing "FTP (File Transfer Protocol)" as the selected protocol. The "Server:" field is empty, and the "Port:" is set to "21". There are also fields for "URL:", "Username:", and "Password:". At the bottom of the dialog are "Connect" and "Cancel" buttons, and a "More Options" dropdown menu.

The Windows taskbar at the bottom shows various application icons including Internet Explorer, Chrome, File Explorer, Word, Excel, PowerPoint, and a yellow duck icon. The system tray on the right shows the date and time as "FR 11:19 AM 7/31/2015".

Data import using FTP

The screenshot displays the Galaxy/ABiMS web interface. The main content area is titled "Download data directly from web or upload files from your disk" and includes a "Get Data" section with a "COMMON TOOLS" subsection. A modal dialog box is open, providing instructions: "allows you to upload files via FTP. To upload some files, log in at **galaxy.sb-roscoff.fr** using your Galaxy credentials (username and password)." Below this, a message states "Your FTP directory does not contain any files." At the bottom of the dialog, there are buttons for "local file", "Choose FTP file", "Paste/Fetch data", "Start", "Pause", "Reset", and "Close".

Overlaid on the interface is the Cyberduck application window, which is in the "Open Connection" state. The connection type is set to "FTP (File Transfer Protocol)". The server is "galaxy.sb-roscoff.fr" and the port is "21". The URL is "ftp://lecorguille@galaxy.sb-roscoff.fr:21/". The username is "lecorguille" and the password is masked with dots. There are checkboxes for "Anonymous Login" and "Save Password", both of which are currently unchecked. The "Connect" button is highlighted by the mouse cursor.

The desktop environment includes a taskbar with icons for Internet Explorer, Google Chrome, File Explorer, Microsoft Word, Microsoft Excel, Microsoft PowerPoint, a calculator, a rubber duck, and a traffic cone. The system tray shows the date and time as "FR 11:19 AM 7/31/2015".

Data import using FTP

The screenshot displays the Galaxy/ABiMS web interface. The main content area is titled "Download data directly from web or upload files from your disk" and includes a "You can Drag & Drop files into this box." area. A sidebar on the left lists "Tools" and "Get Data" options. A modal dialog box is open, titled "Unsecured FTP connection", with the text: "Password will be sent in plaintext. Please contact your web hosting service provider for assistance." The dialog has "Continue" and "Disconnect" buttons, and a "Don't show again" checkbox. Below the dialog, a text box states "Your FTP directory does not contain any files." and a row of buttons includes "local file", "Choose FTP file", "Paste/Fetch data", "Start", "Pause", "Reset", and "Close". The desktop background shows a Windows 7 environment with a taskbar containing icons for Internet Explorer, Chrome, File Explorer, Word, Excel, PowerPoint, and a duck icon. The system tray shows the date and time as 11:19 AM on 7/31/2015.

Data import using FTP

The screenshot displays the Galaxy/ABiMS web interface with a central panel titled "Download data directly from web or upload files from your disk". This panel includes a "You can Drag & Drop files into this box." area and a text box explaining that users can upload files via FTP by logging in at galaxy.sb-roscoff.fr. Below this, a message states "Your FTP directory does not contain any files." At the bottom of the panel are buttons for "local file", "Choose FTP file", "Paste/Fetch data", "Start", "Pause", "Reset", and "Cancel".

An FTP client window titled "lecorguille@galaxy.sb-roscoff.fr - FTP" is overlaid on the left. It shows a menu bar (File, Edit, View, Go, Bookmark, Window, Help), a "Quick Connect" dropdown, and a search bar. The main area contains a table with columns for "Filename", "Size", and "Modified", which is currently empty. A "Copy to /" button is visible at the bottom left of the client window.

A system notification in the bottom right corner reads "Connection opened galaxy.sb-roscoff.fr". The Windows taskbar at the bottom shows various application icons and the system clock indicating 11:19 AM on 7/31/2015.

Data import using FTP

The screenshot displays a Windows desktop environment with a web browser window open to the Galaxy web interface at `galaxy.sb-roscoff.fr`. The browser window shows a "Transfers" dialog box in the foreground, indicating the upload of a file named `left_kept_reads.bam`. The progress bar shows that 50.6 MiB (53,018,624 bytes) of 91.6 MiB (55%) has been transferred at a rate of 70.1 MB/sec, with 1 second remaining. The local file path is `C:\Users\lecorguille\Desktop\left_kept_reads.bam` and the remote URL is `ftp://galaxy.sb-roscoff.fr/left_kept_reads.bam`. The background shows the Galaxy web interface with a "Drop files into this box" area and a message stating "Drop files via FTP. To upload some files, log off.fr using your Galaxy credentials". A system tray notification in the bottom right corner indicates "Connection opened galaxy.sb-roscoff.fr".

Data import using FTP

The screenshot shows the Galaxy/ABiMS web interface with a modal window for data upload. The modal has the title "Download data directly from web or upload files from your disk" and contains the text "You can Drag & Drop files into this box." Below this is a large dashed box for file upload. A yellow message box states "Your FTP directory does not contain any files." At the bottom of the modal are buttons for "local file", "Choose FTP file", "Paste/Fetch data", "Start", "Pause", "Reset", and "Cancel".

An FTP client window titled "lecorguille@galaxy.sb-roscoff.fr - FTP" is open in the foreground. It shows a file list with the following details:

Filename	Size	Modified
left_kept_reads.bam	91.6 MiB	7/31/2015 9:19:00 AM

The desktop taskbar at the bottom shows various application icons including Internet Explorer, Chrome, File Explorer, Word, Excel, PowerPoint, and a yellow duck icon. The system tray on the right shows the date and time as 11:19 AM on 7/31/2015. A notification bubble in the bottom right corner says "Upload complete left_kept_reads.bam".

Data import using FTP

The screenshot shows the Galaxy/ABiMS web interface. The browser address bar displays 'galaxy.sb-roscoff.fr'. The main navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. A 'Tools' sidebar is visible on the left. A large dialog box is open in the center, titled 'Download data directly from web or upload files from your disk'. Inside this dialog, there is a dashed box for file upload and a smaller 'FTP files' sub-dialog. The 'FTP files' dialog contains the text: 'This Galaxy server allows you to upload files via FTP. To upload some files, log in to the FTP server at **galaxy.sb-roscoff.fr** using your Galaxy credentials (email address and password).'. Below this text is a yellow warning box that says 'Your FTP directory does not contain any files.'. At the bottom of the main dialog, there are buttons for 'Choose local file', 'Choose FTP file', 'Paste/Fetch data', 'Start', 'Pause', 'Reset', and 'Close'. A mouse cursor is hovering over the 'Choose FTP file' button. In the bottom right corner, a notification bubble says 'Upload complete'.

Data import using FTP

The screenshot shows a web browser window with the URL `galaxy.sb-roscoff.fr`. The page title is "Galaxy / ABiMS". The navigation menu includes "Analyze Data", "Workflow", "Shared Data", "Visualization", "Admin", "Help", and "User". A modal dialog box is open with the title "Download data directly from web or upload files from your disk". The dialog contains a large dashed box for file upload with the text "You can Drag & Drop files into this box." Below the box are two dropdown menus: "Type (set all):" set to "Auto-detect" and "Genome (set all):" set to "unspecified (?)". At the bottom of the dialog are buttons for "Choose local file", "Choose FTP file", "Paste/Fetch data", "Start", "Pause", "Reset", and "Close". The background interface shows a sidebar with tool categories like "COMMON TOOLS", "SEARCHING TOOLS", and "NGS TOOLS". The Windows taskbar at the bottom shows the time as 11:19 AM on 7/31/2015.

Data import using FTP

The screenshot shows the Galaxy/ABiMS web interface. A dialog box titled "Download data directly from web or upload files from your disk" is open. Inside the dialog, there is a section for "FTP files" which includes instructions on how to upload files via FTP and a table of available files. The table lists one file: "left_kept_reads.bam" with a size of 96 MB and a creation date of 07/31/2015 11:19:45 AM. The dialog also features buttons for "Choose local file", "Choose FTP file", "Paste/Fetch data", "Start", "Pause", "Reset", and "Close".

Galaxy / ABiMS

Analyze Data Workflow Shared Data Visualization Admin Help User

Tools

search tools

Get Data

COMMON TOOLS

Convert Formats

FASTA manipulation

Filter and Sort

Join, Subtract and

Text Manipulation

Graphics

Statistics

EMBOSS 5 Suite

SEARCHING TOOLS

NCBI BLAST+

Diamond

Primer/Microsatelli

NGS TOOLS

NGS: BedTools

NGS: Mapping

NGS: Picard

NGS: QC and manip

Download data directly from web or upload files from your disk

You can Drag & Drop files into this box.

FTP files

This Galaxy server allows you to upload files via FTP. To upload some files, log in to the FTP server at **galaxy.sb-roscoff.fr** using your Galaxy credentials (email address and password).

Available files: 1 files 96 MB

<input type="checkbox"/>	Name	Size	Created
<input type="checkbox"/>	left_kept_reads.bam	96 MB	07/31/2015 11:19:45 AM

Type (set a) specified (?)

Choose local file Choose FTP file Paste/Fetch data Start Pause Reset Close

lecorguille

left_kept_r...

Using 46%

Assets

History

Library is empty. You can upload data or get data from an external source

ABiMS

FR 11:19 AM 7/31/2015

Data import using FTP

Galaxy / ABiMS

galaxy.sb-roscoff.fr

Download data directly from web or upload files from your disk

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

Name	Size	Type	Genome	Settings	Status
left_kept_reads.bam	96 MB	Auto-det...	unspecified (?)		

FTP files

This Galaxy server allows you to upload files via FTP. To upload some files, log in to the FTP server at **galaxy.sb-roscoff.fr** using your Galaxy credentials (email address and password).

Available files: 1 files 96 MB

<input checked="" type="checkbox"/>	Name	Size	Created
<input checked="" type="checkbox"/>	left_kept_reads.bam	96 MB	07/31/2015 11:19:45 AM

Type (set a...)

Choose local file | **Choose FTP file** | Paste/Fetch data | Start | Pause | Reset | Close

Using 46%

Assets

History

Library is empty. You can download data or get data from an external source

FR 11:19 AM 7/31/2015

Data import using FTP

Galaxy / ABiMS

galaxy.sb-roscoff.fr

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 46%

Download data directly from web or upload files from your disk

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

Name	Size	Type	Genome	Settings	Status
left_kept_reads.bam	96 MB	Auto-det...	unspecified (?)	⚙	🗑

Type (set all): Auto-detect Genome (set all): unspecified (?)

Choose local file Choose FTP file Paste/Fetch data **Start** Pause Reset Close

lecorguille

left_kept_r...

COMMON TOOLS

Convert Formats

FASTA manipulation

Filter and Sort

Join, Subtract and C

Text Manipulation

Graphics

Statistics

EMBOSS 5 Suite

SEARCHING TOOLS

NCBI BLAST+

Diamond

Primer/Microsatelli

NGS TOOLS

NGS: BedTools

NGS: Mapping

NGS: Picard

NGS: QC and manip

ABiMS

FR 11:20 AM 7/31/2015

Data import using FTP

Galaxy / ABiMS

galaxy.sb-roscoff.fr

Galaxy / ABiMS

Analyze Data Workflow Shared Data Visualization Admin Help User

Tools

search tools

Get Data

COMMON TOOLS

Convert Formats

FASTA manipulation

Filter and Sort

Join, Subtract and C

Text Manipulation

Graphics

Statistics

EMBOSS 5 Suite

SEARCHING TOOLS

NCBI BLAST+

Diamond

Primer/Microsatelli

NGS TOOLS

NGS: BedTools

NGS: Mapping

NGS: Picard

NGS: QC and manip

Download data directly from web or upload files from your disk

You can Drag & Drop files into this box.

Name	Size	Type	Genome	Settings	Status
left_kept_reads.bam	96 MB	Auto-det...	unspecified (?)		100%

Type (set all): Auto-detect

Genome (set all): unspecified (?)

Choose local file Choose FTP file Paste/Fetch data Start Pause Reset Close

Using 46%

assets

actory

ads.bam

FR 11:20 AM 7/31/2015

Data import using FTP

The screenshot displays the Galaxy / ABiMS web interface in a browser window. The address bar shows the URL `galaxy.sb-roscoff.fr`. The page features a navigation menu with categories like "Analyze Data", "Workflow", "Shared Data", "Visualization", "Admin", "Help", and "User". A green banner at the top reads "Welcome to galaxy.sb-roscoff.fr". Below it, an information box provides contact details for support. The main content area includes the ABiMS logo and the text "Analyses and Bioinformatics for Marine Science", along with links to "Changelog" and "Tutorials". A paragraph describes Galaxy as an open, web-based platform for data-intensive biomedical research. On the left, a sidebar lists various tool categories such as "Get Data", "COMMON TOOLS", "FASTA manipulation", and "NGS TOOLS". The right sidebar shows a "History" section with a search bar and a list of datasets, including "left_kept_reads.bam". The Windows taskbar at the bottom shows the system tray with the date and time "11:20 AM 7/31/2015".

Data import using FTP

The screenshot displays the Galaxy / ABiMS web interface. The browser address bar shows `galaxy.sb-roscoff.fr`. The main navigation bar includes links for **Analyze Data**, **Workflow**, **Shared Data**, **Visualization**, **Admin**, **Help**, and **User**. A green banner at the top reads "Welcome to galaxy.sb-roscoff.fr". Below it, an information box provides contact details: "For any question or request for tools or account, send an email at support.abims@sb-roscoff.fr".

The central content area features the **ABiMS** logo (Analyses and Bioinformatics for Marine Science) and the **Station Biologique Roscoff** logo. Below the logo are expandable sections for **Changelog** and **Tutorials**. A paragraph of text describes the Galaxy platform: "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."

The left sidebar contains a **Tools** section with a search bar and a list of tool categories: **Get Data**, **COMMON TOOLS**, **Convert Formats**, **FASTA manipulation**, **Filter and Sort**, **Join, Subtract and Group**, **Text Manipulation**, **Graphics**, **Statistics**, **EMBOSS 5 Suite**, **SEARCHING TOOLS**, **NCBI BLAST+**, **Diamond**, **Primer/Microsatellite**, **NGS TOOLS**, **NGS: BedTools**, **NGS: Mapping**, **NGS: Picard**, and **NGS: QC and manipulation**.

The right sidebar shows the **History** section with a search bar and a list of datasets. The current history entry is **left_kept_reads.bam**, which is 0 bytes in size.

The Windows taskbar at the bottom shows various application icons, including Internet Explorer, Google Chrome, Word, Excel, PowerPoint, and a duck icon. The system tray on the right indicates the date and time: **FR 11:20 AM 7/31/2015**.

Data import using FTP

The screenshot displays the Galaxy / ABiMS web interface. The browser address bar shows `galaxy.sb-roscoff.fr`. The main navigation bar includes **Galaxy / ABiMS** and menu items: **Analyze Data**, **Workflow**, **Shared Data**, **Visualization**, **Admin**, **Help**, and **User**. A status indicator in the top right corner shows **Using 46%**.

The left sidebar contains a **Tools** section with a search bar and a list of tool categories: **Get Data**, **COMMON TOOLS**, **Convert Formats**, **FASTA manipulation**, **Filter and Sort**, **Join, Subtract and Group**, **Text Manipulation**, **Graphics**, **Statistics**, **EMBOSS 5 Suite**, **SEARCHING TOOLS**, **NCBI BLAST+**, **Diamond**, **Primer/Microsatellite**, **NGS TOOLS**, **NGS: BedTools**, **NGS: Mapping**, **NGS: Picard**, and **NGS: QC and manipulation**.

The main content area features a green banner: **Welcome to galaxy.sb-roscoff.fr**. Below it is an **Information** box with a contact email: `support.abims@sb-roscoff.fr`. The center of the page displays the **ABiMS** logo (Analyses and Bioinformatics for Marine Science) and the **Station Biologique Roscoff** logo. Navigation links for **Changelog** and **Tutorials** are provided. A paragraph of text describes the Galaxy platform and its affiliations.

The right sidebar shows the **History** section with a search bar and **Unnamed history** (1 shown). A file entry is visible: **left_kept_reads.bam** (91.6 MB, format: bam, database: ?). The file is marked as an **uploaded bam file** and includes options to **display in IGB View** and **Binary bam alignments file**.

The Windows taskbar at the bottom shows the system tray with the date **7/31/2015** and time **11:20 AM**, along with various application icons.

~~DATA IMPORT~~

Data import

For HUGE public resources: genome, databank ...

--> Make a request to the support team

The screenshot displays the Galaxy/ABiMS interface for the NCBI BLAST+ tool. The main panel shows the tool configuration for a BLAST search. The 'Nucleotide query sequence(s)' field is empty, showing 'No fasta dataset available.' The 'Subject database/sequences' is set to 'Locally installed BLAST database'. A red box highlights the 'Nucleotide BLAST database' section, which includes a 'Select/Unselect all' checkbox and a list of databases: nt, genbank, genbank Bacterial, genbank Environmental sampling, genbank EST (expressed sequence tag), genbank GSS (genome survey sequence), genbank HTC (high throughput cDNA sequencing), and genbank HTGS (high throughput genomic sequencing). The 'Set expectation value cutoff' is set to 0.001. The 'Output format' is set to 'Tabular (with 105 columns)'. The left sidebar shows various tool categories, and the right sidebar shows a history of datasets, including 'eba 2016 sartools' and several 'SARTools DESeq2' outputs.

Hands-on

DATA IMPORT





1. Fetch the files with your internet browser:
<https://tinyurl.com/training-galaxy-initiation>

2. Upload files into Galaxy

- a. From URL: <https://tinyurl.com/exons-file>
- b. From disk: snps.bed
- c. Using FTP: repeats.bed

TOOLS

Tools - panel

The screenshot displays the Galaxy / ABiMS web interface. At the top, a navigation bar includes 'Galaxy / ABiMS', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', 'User', and a 'Using 0%' indicator. The left sidebar, titled 'Tools', contains a search bar and a list of tool categories: 'Get Data', 'Send Data', 'Collection Operations', 'COMMON TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Convert Formats', 'Extract Features', 'Fetch Sequences', 'Statistics', 'Graph/Display Data', 'Fasta Fastq Manipulation', and 'COMMON NGS TOOLS'. The 'Fasta Fastq Manipulation' section is expanded, showing options like 'Filter sequences by ID from a tabular file', 'FastQC Read Quality reports', and 'FASTQ Groomer'. The main content area features a green 'Welcome to galaxy3.sb-roscoff.fr' message, an information box with contact details for support.abims@sb-roscoff.fr, the ABiMS logo (Analyses and Bioinformatics for Marine Science), and links to 'Changelog' and 'Tutorials'. A paragraph of text describes Galaxy as an open, web-based platform for data-intensive biomedical research, mentioning its affiliation with Penn State and Emory University. The right sidebar, titled 'History', shows a search bar and a list of datasets: 'Trinity example' (2 shown, 3 deleted, 40.02 KB) and two files named '4: reads.left.fg' and '3: reads.right.fg'.

Tools - panel

The screenshot displays the Galaxy / ABiMS web interface. At the top, the navigation bar includes 'Galaxy / ABiMS', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. A 'Using 0%' indicator is visible in the top right corner.

The left sidebar, titled 'Tools', contains a search bar and several categories of tools:

- Get Data**
- Send Data**
- Collection Operations**
- COMMON TOOLS**
 - Text Manipulation**
 - Filter and Sort**
 - Join, Subtract and Group**
 - Convert Formats**
 - Extract Features**
 - Fetch Sequences**
 - Statistics**
 - Graph/Display Data**
 - Fasta Fastq Manipulation**
 - Filter sequences by ID from a tabular file
 - FastQC Read Quality reports
 - FASTQ Groomer convert between various FASTQ quality formats
- COMMON NGS TOOLS**
 - NGS:Samtools**

The central workspace features a green 'Welcome to galaxy3.sb-roscoff.fr' message, an information box with contact details for support.abims@sb-roscoff.fr, and the ABiMS logo (Analyses and Bioinformatics for Marine Science) with the Station Biologique Roscoff emblem.

The right sidebar, titled 'History', shows a search bar for datasets and a list of recent jobs:

- Trinity example**: 2 shown, 3 deleted, 40.02 KB
- 4: reads.left.fg**
- 3: reads.right.fg**

A large white box with a black border is overlaid on the central workspace, containing the text: 'What tools are available?'

Below the box, a paragraph of text describes Galaxy: '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.'

125+ platforms for using Galaxy: <https://galaxyproject.org/use>
[Servers, clouds, deployable resources]

125+ platforms for using Galaxy: <https://galaxyproject.org/use>
[Servers, clouds, deployable resources]

GENERAL PURPOSE:

UseGalaxy servers:

usegalaxy.org, usegalaxy.eu, usegalaxy.org.au, usegalaxy.fr (coming)

125+ platforms for using Galaxy: <https://galaxyproject.org/use>
[Servers, clouds, deployable resources]

GENERAL PURPOSE:

Use Galaxy servers:

usegalaxy.org, usegalaxy.eu, usegalaxy.org.au, usegalaxy.fr (coming)

DOMAIN SPECIFIC:



RNAseq: <http://galaxy.sb-roscoff.fr>

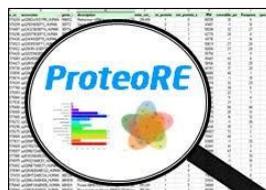
SBR tools: <http://webtools.sb-roscoff.fr>

Metagenomics: <http://galaxy4metab.sb-roscoff.fr>

Metabolomics:



Proteomics:



ChIP-seq:



125+ platforms for using Galaxy: <https://galaxyproject.org/use>
 [Servers, clouds, deployable resources]

GENERAL PURPOSE:

Use Galaxy servers:

usegalaxy.org, usegalaxy.eu, usegalaxy.org.au, usegalaxy.fr (coming)

DOMAIN SPECIFIC:



RNAseq: <http://galaxy.sb-roscoff.fr>

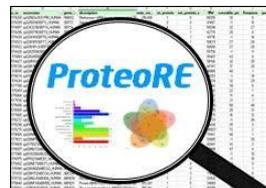
SBR tools: <http://webtools.sb-roscoff.fr>

Metagenomics: <http://galaxy4metab.sb-roscoff.fr>

Metabolomics:



Proteomics:



ChIP-seq:



Catalog of French Galaxy tools: <http://galaxycat.france-bioinformatique.fr>

Tools - panel

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools

trinity

Trinity suite

1- ASSEMBLY:

[Trinity](#) de novo assembly of RNA-Seq data

[Trinity Statistics](#) Obtain basic stats for the number of genes and isoforms and contiguity of the assembly

[Generate gene to transcript map](#) for Trinity assembly

2- COUNTING:

[Align reads and estimate abundance](#) on a de novo assembly of RNA-Seq data

[Build expression matrix](#) for a de novo assembly of RNA-Seq data by Trinity

3- DIFFERENTIAL EXPRESSION:

[RNASeq samples quality check](#) for transcript quantification

✓ Welcome to galaxy3.sb-roscoff.fr

Information
For any question or request for tools or account, send an email at support.abims@sb-roscoff.fr

ABiMS
Analyses and Bioinformatics for Marine Science

Station Biologique Roscoff

Changelog

Tutorials

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

search datasets

Trinity example
2 shown, 3 deleted

40.02 KB

4: reads.left.fq

3: reads.right.fq

Tools - form

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools

trinity

Trinity suite

1- ASSEMBLY:

[Trinity de novo assembly of RNA-Seq data](#)

[Trinity Statistics](#) Obtain basic stats for the number of genes and isoforms and contiguity of the assembly

[Generate gene to transcript map](#) for Trinity assembly

2- COUNTING:

[Align reads and estimate abundance](#) on a de novo assembly of RNA-Seq data

[Build expression matrix](#) for a de novo assembly of RNA-Seq data by Trinity

3- DIFFERENTIAL EXPRESSION:

[RNASeq samples quality check](#) for transcript quantification

Trinity de novo assembly of RNA-Seq data (Galaxy Version 2.2.0.0) Options

Paired or Single-end data?

Single

Single-end reads

4: reads.left.fq
3: reads.right.fq

(--single)

Strand specific data

Yes No

Run in silico normalization of reads

Yes No

Defaults to max. read coverage of 50. (--normalize_reads)

[Additional Options](#)

Execute

[Trinity](#) assembles transcript sequences from Illumina RNA-Seq data.

Citations [Show BibTeX](#)

Grabherr, Manfred G and Haas, Brian J and Yassour, Moran and Levin, Joshua Z and Thompson, Dawn A and Amit, Ido and Adiconis, Xian and Fan, Lin and Raychowdhury, Raktima and Zeng, Qiandong and et al. (2011). Full length

History

search datasets

Trinity example
2 shown, 2 [deleted](#)

37.53 KB

4: reads.left.fq

3: reads.right.fq

Tools - form

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools

trinity

Trinity suite

1- ASSEMBLY:

[Trinity de novo assembly of RNA-Seq data](#)

[Trinity Statistics](#) Obtain basic stats for the number of genes and isoforms and contiguity of the assembly

[Generate gene to transcript map](#) for Trinity assembly

2- COUNTING:

[Align reads and estimate abundance](#) on a de novo assembly of RNA-Seq data

[Build expression matrix](#) for a de novo assembly of RNA-Seq data by Trinity

3- DIFFERENTIAL EXPRESSION:

[RNASeq samples quality check](#) for transcript quantification

Trinity de novo assembly of RNA-Seq data (Galaxy Version 2.2.0.0) Options

Paired or Single-end data?

Single

Single-end reads

4: reads.left.fq
3: reads.right.fq

(--single)

Strand specific data

Yes No

Run in silico normalization of reads

Yes No

Defaults to max. read coverage of 50. (--normalize_reads)

Additional Options

Execute

Trinity assembles transcript sequences from Illumina RNA-Seq data.

Citations [Show BibTeX](#)

Grabherr, Manfred G and Haas, Brian J and Yassour, Moran and Levin, Joshua Z and Thompson, Dawn A and Amit, Ido and Adiconis, Xian and Fan, Lin and Raychowdhury, Raktima and Zeng, Qiandong and et al. (2011). Full length

History

search datasets

Trinity example

2 shown, 2 [deleted](#)

37.53 KB

4: reads.left.fq

3: reads.right.fq

Tools - form

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools

trinity

Trinity suite

1- ASSEMBLY:

[Trinity de novo assembly of RNA-Seq data](#)

[Trinity Statistics](#) Obtain basic stats for the number of genes and isoforms and contiguity of the assembly

[Generate gene to transcript map](#) for Trinity assembly

2- COUNTING:

[Align reads and estimate abundance](#) on a de novo assembly of RNA-Seq data

[Build expression matrix](#) for a de novo assembly of RNA-Seq data by Trinity

3- DIFFERENTIAL EXPRESSION:

[RNASeq samples quality check](#) for transcript quantification

Trinity de novo assembly of RNA-Seq data (Galaxy Version 2.2.0.0) Options

Paired or Single-end data?

Single

Single-end reads

4: reads.left.fq
3: reads.right.fq

(--single)

Strand specific data

Yes No

Run in silico normalization of reads

Yes No

Defaults to max. read coverage of 50. (--normalize_reads)

Additional Options

Execute

[Trinity](#) assembles transcript sequences from Illumina RNA-Seq data.

Citations [Show BibTeX](#)

Grabherr, Manfred G and Haas, Brian J and Yassour, Moran and Levin, Joshua Z and Thompson, Dawn A and Amit, Ido and Adiconis, Xian and Fan, Lin and Raychowdhury, Raktima and Zeng, Qiandong and et al. (2011). Full length

History

search datasets

Trinity example

2 shown, 2 [deleted](#)

37.53 KB

4: reads.left.fq

3: reads.right.fq

Tools - form

Tools can have some advanced options

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools

trinity

Trinity suite

1- ASSEMBLY:

[Trinity](#) de novo assembly of RNA-Seq data

[Trinity Statistics](#) Obtain basic stats for the number of genes and isoforms and contiguity of the assembly

[Generate gene to transcript map](#) for Trinity assembly

2- COUNTING:

[Align reads and estimate abundance](#) on a de novo assembly of RNA-Seq data

[Build expression matrix](#) for a de novo assembly of RNA-Seq data by Trinity

3- DIFFERENTIAL EXPRESSION:

[RNASeq samples quality check](#) for transcript quantification

Run in silico normalization of reads

Yes No

Defaults to max. read coverage of 50. (--normalize_reads)

Additional Options

Minimum Contig Length

200

All contigs shorter than this will be discarded (--min_contig_length)

Use the genome guided mode?

No

If you already mapped the reads to the genome, Trinity can use this information

Error-corrected or circular consensus (CCS) pac bio reads

Nothing selected

Experimental feature! Long reads must be in the same orientation as short reads if they are strand specific (--long_reads)

Minimum count for K-mers to be assembled

1

(--min_kmer_cov)

Execute

[Trinity](#) assembles transcript sequences from Illumina RNA-Seq data.

Citations Show BibTeX

History

search datasets

Trinity example

2 shown, 2 deleted

37.53 KB

4: reads.left.fq

3: reads.right.fq

Tools - form

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools

trinity

Trinity suite

1- ASSEMBLY:

[Trinity](#) de novo assembly of RNA-Seq data

[Trinity Statistics](#) Obtain basic stats for the number of genes and isoforms and contiguity of the assembly

[Generate gene to transcript map](#) for Trinity assembly

2- COUNTING:

[Align reads and estimate abundance](#) on a de novo assembly of RNA-Seq data

[Build expression matrix](#) for a de novo assembly of RNA-Seq data by Trinity

3- DIFFERENTIAL EXPRESSION:

[RNASeq samples quality check](#) for transcript quantification

Trinity de novo assembly of RNA-Seq data (Galaxy Version 2.2.0.0) Options

Paired or Single-end data?

Paired

Left/Forward strand reads

4: reads.left.fq
3: reads.right.fq

(--left)

Right/Reverse strand reads

4: reads.left.fq
3: reads.right.fq

(--right)

Strand specific data

Yes No

Run in silico normalization of reads

Yes No

Defaults to max. read coverage of 50. (--normalize_reads)

[Additional Options](#)

Execute

History

search datasets

Trinity example
2 shown, 3 deleted

40.02 KB

4: reads.left.fq

3: reads.right.fq

Tools - form

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools

trinity

Trinity suite

1- ASSEMBLY:

[Trinity](#) de novo assembly of RNA-Seq data

[Trinity Statistics](#) Obtain basic stats for the number of genes and isoforms and contiguity of the assembly

[Generate gene to transcript map](#) for Trinity assembly

2- COUNTING:

[Align reads and estimate abundance](#) on a de novo assembly of RNA-Seq data

[Build expression matrix](#) for a de novo assembly of RNA-Seq data by Trinity

3- DIFFERENTIAL EXPRESSION:

[RNASeq samples quality check](#) for transcript quantification

1 job has been successfully added to the queue - resulting in the following datasets:

5: Trinity on data 3 and data 4: Assembled Transcripts

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History

search datasets

Trinity example
3 shown, 2 deleted

37.53 KB

5: Trinity on data 3 and data 4: Assembled Transcripts

4: reads.left.fq

3: reads.right.fq

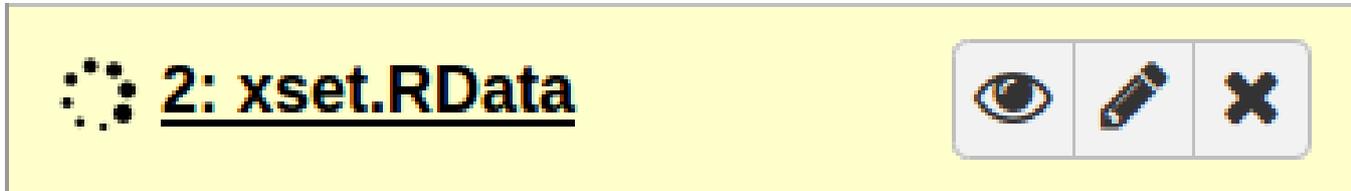
 2: xset.RData



Job is waiting to run

= the job is in the scheduler « queue »

Duration time of this status depends on the amount of actual queued jobs and on the requested number of processors



Job is currently running

= the job is being executed on the computing cluster

Duration time of this status depends on the job's attributes and the computing resources allocated.

Some programs are executed with several processors (using 4, 8 or 16 Gb of RAM).

And others are mono-threaded ☹️

2: xset.RData



Job is finished and status is OK

But warnings or errors can be hidden behind!



16: xset.RData



Job is finished but with an error status

= the program sends an error

The error is often explained by the program but sometimes ...
not.



16: xset.RData



Job is finished but with an error status

= the program sends an error

Possible causes of error :

- ~~The user~~ :P
- Bad usage : input file, format or option
- Bad integration of the program into Galaxy ... sorry :/
- Non anticipated crash of the program

Tools - Handle errors

The screenshot displays the Galaxy / ABiMS web interface. At the top, the navigation bar includes 'Galaxy / ABiMS', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', 'User', and a 'Using 0%' indicator. The left sidebar contains a 'Tools' section with a search bar and a list of categories: 'Get Data', 'Send Data', 'Collection Operations', 'COMMON TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Convert Formats', 'Extract Features', 'Fetch Sequences', 'Statistics', 'Graph/Display Data', 'Fasta Fastq Manipulation', 'COMMON NGS TOOLS', 'NGS:Samtools', 'NGS:Mapping', 'NGS:Bedtools', 'NGS:Picard Tools', 'SEARCHING TOOLS', and 'Diamond'. The main content area features a green 'Welcome to galaxy3.sb-roscoff.fr' message, an 'Information' box with contact details for support.abims@sb-roscoff.fr, the ABiMS logo (Analyses and Bioinformatics for Marine Science), and links to 'Changelog' and 'Tutorials'. Below these is a paragraph describing Galaxy as an open, web-based platform for data-intensive biomedical research, supported by various institutions. The right sidebar shows a 'History' panel with a search bar and a list of datasets, including 'eba 2016 sartools', 'group2_count2.txt', 'group2_count1.txt', 'group1_count2.txt', 'group1_count1.txt', 'SARTools edgeR R objects (.RData)', 'SARTools edgeR R log', 'SARTools edgeR figures', 'SARTools edgeR tables', and 'SARTools edgeR'.

Tools - Handle errors

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools search tools

Get Data
Send Data
Collection Operations

COMMON TOOLS
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Statistics
Graph/Display Data
Fasta Fastq Manipulation

COMMON NGS TOOLS
NGS:Samtools
NGS:Mapping
NGS:Bedtools
NGS:Picard Tools

SEARCHING TOOLS
Diamond

Dataset generation errors

Dataset 48: group2_count2.txt

The Galaxy framework encountered the following error while attempting to run the tool:

```
Traceback (most recent call last):  
  File "/w/galaxy/galaxy3/galaxy/lib/galaxy/jobs/runners/local.py", line  
    stdout_file.close()  
  File "/opt/python/lib/python2.7/tempfile.py", line 403, in close  
    self.unlink(self.name)  
OSError: [Errno 2] No such file or directory: '/w/galaxy/galaxy3/galaxy/database'
```

Tool execution generated the following error message:

```
failure running job
```

Report this error to the local Galaxy administrators

Usually the local Galaxy administrators regularly review errors that occur on the server. 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

galaxy3.sb-roscoff.fr/dataset/errors?id=53eb0ef5c8056a28 e.gueguen@sb-roscoff.fr

History search datasets

eba 2016 sartools
28 shown, 14 deleted
1.59 MB

48: group2_count2.txt
tool error
An error occurred with this dataset:
failure running job

View or report this error

gene1	1353
gene10	72
gene100	496
gene1000	50

47: group2_count1.txt
46: group1_count2.txt
45: group1_count1.txt

Tools - Handle errors

Sent to the support team

The screenshot shows the Galaxy/ABiMS web interface. At the top, there is a navigation bar with 'Galaxy / ABiMS' and various menu items like 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. A 'Using 0%' indicator is visible in the top right.

The main content area is titled 'Tool execution generated the following error message:'. Below this, a dashed box contains the text 'failure running job'. A large heading reads 'Report this error to the local Galaxy administrators'. The text below explains that local administrators review errors and that providing additional information and an email address will help them investigate.

An 'Error Report' form is displayed with the following fields:

- Your email:** A text input field containing 'loraine.gueguen@sb-roscoff.fr'.
- Message:** A large text area for providing details about the error.
- Report:** A button to submit the error report.

On the right side, a 'History' panel is visible, showing a list of datasets. The top dataset is 'eba 2016 sartools' (1.59 MB). Below it, a red error entry is highlighted, labeled '48: group2_count2.txt', with the message 'tool error: An error occurred with this dataset: failure running job'. Below the error entry is a table of gene counts:

Gene	Count
1.gene0	2,1813
gene1	1353
gene10	72
gene100	496
gene1000	50

Below the error entry are other dataset entries: '47: group2_count1.txt', '46: group1_count2.txt', and '45: group1_count1.txt'. A red box highlights a scroll bar on the right side of the history panel.

DATASET

Dataset

Both inputs and outputs

The screenshot displays the Galaxy / ABiMS web interface. The top navigation bar includes 'Galaxy / ABiMS', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', 'User', and a 'Using 0%' indicator. The left sidebar contains a 'tools' section with a search bar and various tool categories like 'Get Data', 'Send Data', 'Collection Operations', 'COMMON TOOLS', 'COMMON NGS TOOLS', and 'SEARCHING TOOLS'. The main content area features a green welcome banner for 'galaxy3.sb-roscoff.fr', an information box with contact details, and the ABiMS logo with the text 'Analyses and Bioinformatics for Marine Science'. Below the logo are links for 'Changelog' and 'Tutorials'. A paragraph of text describes the Galaxy platform and its affiliations. On the right, the 'History' panel shows a list of datasets: 'Trinity example' (40.3 KB), '5: Trinity on data 3 and data 4: Assembled Transcripts' (highlighted in green and red), '4: reads.left.fq', and '3: reads.right.fq'. Each dataset entry includes icons for viewing, editing, and deleting.

Dataset

Informations

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tool: Trinity

Number: 5
Name: Trinity on data 3 and data 4: Assembled Transcripts
Created: Wed 01 Mar 2017 03:52:36 PM (UTC)
Filesize: 2.5 KB
Dbkey: ?
Format: fasta
Galaxy Tool ID: toolshed.g2.bx.psu.edu/repos/iuc/trinity/trinity/2.2.0.0
Galaxy Tool Version: 2.2.0.0
Tool Version:
Tool Standard Output: [stdout](#)
Tool Standard Error: [stderr](#)
Tool Exit Code: 0

Input Parameter	Value	Note for rerun
Paired or Single-end data?	paired	
Left/Forward strand reads	4: reads.left.fq	
Right/Reverse strand reads	3: reads.right.fq	
Strand specific data	true	
Strand-specific Library Type	Reverse-Forward	
Jaccard Clip options	Not available.	
Run in silico normalization of reads	True	
additional_params		
Minimum Contig Length	200	
Use the genome guided mode?	no	

[galaxy3.sb-roscoff.fr/datasets/c10ec933dc50450a/show_params](#)

History search datasets

Trinity example
3 shown, 3 deleted
40.3 KB

5: Trinity on data 3 and data 4: Assembled Transcripts
7 sequences
format: **fasta**, database: ?
[View details](#) [3_g1_i1 len=541 path=\[519:0-544\]](#)
GTCTGAATTGCGATGTAATGCAGCTTTCCAGACACAAGTATGG
TCGCCATTGTGCAAAATATGTGTCTGATAGACCSCAGGCTTTCA
TGACATGAGCGTGGCACCTGAAGACAGGGTGTGGGTGAGAGGGTC
TGAGTTGTCTTGTATCATCAATAGATGCAAAATTAGATGTAAGAAC

4: reads.left.fq

3: reads.right.fq

Dataset

Informations

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tool: Trinity

Number: 5
Name: Trinity on data 3 and data 4: Assembled Transcripts
Created: Wed 01 Mar 2017 03:52:36 PM (UTC)
Filesize: 2.5 KB
Dbkey: ?
Format: fasta
Galaxy Tool ID: toolshed.g2.bx.psu.edu/repos/iuc/trinity/trinity/2.2.0.0
Galaxy Tool Version: 2.2.0.0
Tool Version:
Tool Standard Output: **stdout**
Tool Standard Error: **stderr**
Tool Exit Code: 0

Input Parameter	Value	Note for rerun
Paired or Single-end data?	paired	
Left/Forward strand reads	4: reads.left.fq	
Right/Reverse strand reads	3: reads.right.fq	
Strand specific data	true	
Strand-specific Library Type	Reverse-Forward	
Jaccard Clip options	Not available.	
Run in silico normalization of reads	True	
additional_params		
Minimum Contig Length	200	
Use the genome guided mode?	no	

galaxy3.sb-roscoff.fr/datasets/c10ec933dc50450a/show_params

History search datasets

Trinity example
3 shown, 3 deleted
40.3 KB

5: Trinity on data 3 and data 4: Assembled Transcripts
7 sequences
format: **fasta**, database: ?
View details [3_g1_i1 len=541 path=\[519:0-544\]](#)

```
GTCTGAATTGCGATGTAATGCAGCTTTCCAGACACAAGTATGG  
TCGCCATTGTGCAAAATATGTGTCTGATAGACCSCAGGCTTCA  
TGACATGAGCGTGGCACCTGAAGACAGGGTGTGGGTGAGAGGGTC  
TGAGTTGTCTTGTATCATCAATAGATGCAAAATTAGATGTAAGAAC
```

4: reads.left.fq

3: reads.right.fq

Dataset

Download

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tool: Trinity

Number: 5
Name: Trinity on data 3 and data 4: Assembled Transcripts
Created: Wed 01 Mar 2017 03:52:36 PM (UTC)
Filesize: 2.5 KB
Dbkey: ?
Format: fasta
Galaxy Tool ID: toolshed.g2.bx.psu.edu/repos/iuc/trinity/trinity/2.2.0.0
Galaxy Tool Version: 2.2.0.0
Tool Version:
Tool Standard Output: [stdout](#)
Tool Standard Error: [stderr](#)
Tool Exit Code: 0

Input Parameter	Value	Note for rerun
Paired or Single-end data?	paired	
Left/Forward strand reads	4: reads.left.fq	
Right/Reverse strand reads	3: reads.right.fq	
Strand specific data	true	
Strand-specific Library Type	Reverse-Forward	
Jaccard Clip options	Not available.	
Run in silico normalization of reads	True	
additional_params		
Minimum Contig Length	200	
Use the genome guided mode?	no	

galaxy3.sb-roscoff.fr/datasets/c10ec933dc50450a/display?to_ext=fasta

History search datasets

Trinity example
3 shown, 3 deleted
40.3 KB

5: Trinity on data 3 and data 4: Assembled Transcripts
7 sequences
format: **fasta**, database: ?
[Download](#) [Info](#) [Refresh](#)

```
DN0_c0_g1_i1 len=541 path=[519:0-544]
GTCTGAATTCGCATGTAATGCAGCTTTCCAGACACAAGTATGG
TCGCCATTGTGCAAAATATGTGTCTGATAGACCCGACGGCTTCA
TGACATGAGCGTGGCACCTGAAGACAGGGTGTGGSTGAGAGGGTC
TGAGTTGTCTTGTATCATCAATAGATGCAAAATTAGATGTAAGAAC
```

4: reads.left.fq

3: reads.right.fq

Dataset

Re-run a job

The screenshot displays the Galaxy ABiMS interface for running a Trinity de novo assembly job. The main panel shows the tool configuration for 'Trinity de novo assembly of RNA-Seq data (Galaxy Version 2.2.0.0)'. The 'Paired or Single-end data?' dropdown is set to 'Paired'. Under 'Left/Forward strand reads', the input list contains '5: Trinity on data 3 and data 4: Assembled Transcripts', '4: reads.left.fq' (highlighted), and '3: reads.right.fq'. The 'Right/Reverse strand reads' section has the same list with '3: reads.right.fq' highlighted. The 'Strand specific data' section has 'Yes' selected. The 'Strand-specific Library Type' dropdown is set to 'Reverse-Forward'. The 'Jaccard Clip options' section has 'Yes' selected. The bottom of the tool panel shows the option 'You expect high gene density with UTR overlap (--jaccard_clip)'. The left sidebar contains navigation menus for 'Tools', 'COMMON TOOLS', 'COMMON NGS TOOLS', and 'SEARCHING TOOLS'. The right sidebar shows the 'History' panel with a search bar and a list of datasets. The top dataset is '5: Trinity on data 3 and data 4: Assembled Transcripts' (40.3 KB), which is highlighted in green. Below it are '4: reads.left.fq' and '3: reads.right.fq'. A red circular icon with a refresh symbol is overlaid on the top dataset, with a tooltip that says 'Run this job again'. The bottom of the history panel shows a URL: 'galaxy3.sb-roscoff.fr/tool_runner/rerun?id=c10ec933dc50450a'.

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools search tools

Get Data Send Data Collection Operations COMMON TOOLS Text Manipulation Filter and Sort Join, Subtract and Group Convert Formats Extract Features Fetch Sequences Statistics Graph/Display Data Fasta Fastq Manipulation COMMON NGS TOOLS NGS:Samtools NGS:Mapping NGS:Bedtools NGS:Picard Tools SEARCHING TOOLS Diamond

Trinity de novo assembly of RNA-Seq data (Galaxy Version 2.2.0.0) Options

Paired or Single-end data?
Paired

Left/Forward strand reads
5: Trinity on data 3 and data 4: Assembled Transcripts
4: reads.left.fq
3: reads.right.fq
(--left)

Right/Reverse strand reads
5: Trinity on data 3 and data 4: Assembled Transcripts
4: reads.left.fq
3: reads.right.fq
(--right)

Strand specific data
Yes No

Strand-specific Library Type
Reverse-Forward
(--SS_lib_type)

Jaccard Clip options
Yes No

You expect high gene density with UTR overlap (--jaccard_clip)

History search datasets

Trinity example
3 shown, 3 deleted
40.3 KB

5: Trinity on data 3 and data 4: Assembled Transcripts
7 sequences
format: fasta, database: ?
Run this job again len=541 path=[519:0-541]
GTCTGAATTGGCATGTAATGCAGCTTTCCCGACACACAAGTATGG
TCGCCATTGTGCAAAATATGTGTCTGTATAGACCCGACGGCTTCA
TGACATGAGCGTGGCACCTGAAGACAGGGTGTGGSTGAGAGGGST
TGAGTTGTCTTGATCATCAATAGATGCAAAATTAGATGTAAGBAK

4: reads.left.fq

3: reads.right.fq

galaxy3.sb-roscoff.fr/tool_runner/rerun?id=c10ec933dc50450a

Dataset

Dataset display : text, tabular, pdf, picture, html ...

The screenshot displays the Galaxy/ABiMS web interface. The top navigation bar includes 'Galaxy / ABiMS', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. A 'Using 0%' indicator is visible in the top right. The left sidebar contains a 'Tools' section with a search bar and various tool categories: 'Get Data', 'Send Data', 'Collection Operations', 'COMMON TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Convert Formats', 'Extract Features', 'Fetch Sequences', 'Statistics', 'Graph/Display Data', 'Fasta Fastq Manipulation', 'COMMON NGS TOOLS', 'NGS:Samtools', 'NGS:Mapping', 'NGS:Bedtools', 'NGS:Picard Tools', 'SEARCHING TOOLS', and 'Diamond'. The main area shows a tabular dataset with 6 columns and 20 rows. The columns are labeled 1 through 6, and the rows are labeled with transcript IDs. The right sidebar shows a 'History' panel with a list of operations, including 'cluster differentially expressed transcripts on data 2, data 3, and others', '7: Extract and cluster differentially expressed transcripts on data 2, data 3, and others: extracted differentially expressed genes', '6: de results', '5: matrix.counts.matrix', '4: input.matrix.wt GSNO vs wt ph8.DESeq2.DE results', '3: input.matrix.wt 37 vs wt ph8.DESeq2.DE results', '2: input.matrix.wt 37 vs wt GSNO.DESeq2.DE results', and '1: samples.txt'. A 'View data' tooltip is visible over the '5:' entry. The bottom status bar shows the URL: 'http://galaxy3.sb-roscoff.fr/datasets/4437d546e349a03a/display/?preview=True'.

1	2	3	4	5	6
	wt_37_2	wt_37_3	wt_37_1	wt_GSNO_3	wt_GSNO_1
TR24 c0_g1_i1	90.00	67.00	85.00	36.00	35.00
TR2779 c0_g1_i1	186.00	137.00	217.00	147.00	186.00
TR127 c1_g1_i1	9.00	23.00	16.00	2.00	0.00
TR2107 c1_g1_i1	59.00	65.00	47.00	6.00	6.00
TR2011 c5_g1_i1	11.00	4.00	4.00	8.00	5.00
TR4163 c0_g1_i1	368.00	422.00	425.00	172.00	216.00
TR5055 c0_g2_i1	36.00	17.00	27.00	4.00	7.00
TR1449 c0_g1_i1	196.00	230.00	207.00	66.00	113.00
TR1982 c2_g1_i1	7.00	7.00	6.00	4.00	3.00
TR1859 c3_g1_i1	0.00	0.00	1.00	0.00	0.00
TR1492 c0_g1_i2	1895.00	1906.00	1921.00	1104.00	1263.00
TR1122 c0_g1_i1	2.00	3.00	0.00	3.00	0.00
TR2278 c0_g1_i1	497.00	610.00	598.00	333.00	406.00
TR4084 c0_g1_i1	95.00	148.00	86.00	77.00	111.00
TR4761 c0_g1_i1	2089.00	1746.00	1875.00	155.00	174.00
TR3638 c0_g1_i1	647.00	676.00	712.00	117.00	184.00
TR2090 c0_g1_i1	0.00	0.00	0.00	22.00	0.00
TR3854 c0_g1_i1	1878.00	1734.00	1864.00	1775.00	2173.00
TR131 c0_g1_i1	32.00	28.00	31.00	1001.00	1233.00
TR5075 c0_g1_i1	13.00	22.00	21.00	6.00	8.00
TR2182 c3_g2_i6	1.44	2.70	3.84	3.35	0.00
TR3788 c0_g1_i1	17.00	30.00	22.00	91.00	132.00
TR4859 c0_g1_i1	6.00	12.00	8.00	4.00	1.00
TR2487 c0_g1_i1	386.00	383.00	424.00	689.00	866.00

Dataset

Renaming and annotation

The screenshot displays the Galaxy/ABiMS web interface. The top navigation bar includes 'Galaxy / ABiMS', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. A 'Using 0%' indicator is visible in the top right. The left sidebar contains a search bar and a list of tool categories: 'Get Data', 'Send Data', 'Collection Operations', 'COMMON TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Convert Formats', 'Extract Features', 'Fetch Sequences', 'Statistics', 'Graph/Display Data', 'Fasta Fastq Manipulation', 'COMMON NGS TOOLS', 'NGS:Samtools', 'NGS:Mapping', 'NGS:Bedtools', 'NGS:Picard Tools', 'SEARCHING TOOLS', and 'Diamond'. The main content area is titled 'Edit Attributes' and has four tabs: 'Attributes', 'Convert Format', 'Datatype', and 'Permissions'. The 'Attributes' tab is active, showing the following fields:

- Name:** matrix.counts.matrix
- Info:** uploaded tabular file
- Annotation / Notes:** This is my expression matrix.
- Database/Build:** unspecified (?)

Buttons for 'Save' and 'Auto-detect' are present. A tooltip 'Edit attributes' is visible over the 'matrix.counts.matrix' entry in the history panel. The right sidebar shows the 'History' panel with a list of datasets:

- 8: Extract and cluster differentially expressed transcripts on data 2, data 3, and others
- 7: Extract and cluster differentially expressed transcripts on data 2, data 3, and others: extracted differentially expressed genes
- 6: de results
- 5: matrix.counts.matrix (41 lines, format: txt, database: ?)

The bottom status bar shows the URL: galaxy3.sb-roscoff.fr/datasets/4437d546e349a03a/edit.

Dataset

Renaming and annotation

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools

search tools

Get Data
Send Data
Collection Operations

COMMON TOOLS
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Statistics
Graph/Display Data
Fasta Fastq Manipulation

COMMON NGS TOOLS
NGS:Samtools
NGS:Mapping
NGS:Bedtools
NGS:Picard Tools

SEARCHING TOOLS
Diamond

Attributes Convert Format Datatype Permissions

Edit Attributes

Name:
matrix.counts.matrix

Info:
uploaded tabular file

Annotation / Notes:
This is my expression matrix.
Add an annotation or notes to a dataset; annotations are available when a history is viewed.

Database/Build:
unspecified (?)
Save
Auto-detect
This will inspect the dataset and attempt to correct the above column values if they are not accurate.

History

8: Extract and cluster differentially expressed transcripts on data 2, data 3, and others

7: Extract and cluster differentially expressed transcripts on data 2, data 3, and others: extracted differentially expressed genes
a list of datasets

6: de results
a list of 3 datasets

5: matrix.counts.matrix
41 lines
format: txt, database: ?
uploaded tabular file
Tags: trinity
Annotation: This is my expression matrix.

wt_37_2 wt_37_3 wt_37_1 wt_GSNO_3

galaxy3.sb-roscoff.fr/datasets/4437d546e349a03a/edit

Dataset

Change the Datatype of the Dataset

The screenshot shows the Galaxy / ABiMS web interface. At the top, there are navigation tabs: 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. The 'Datatype' tab is selected and highlighted with a red box. Below the tabs, there is a 'Change data type' section with a 'New Type:' dropdown menu. The dropdown menu is open, showing a list of data types: txt, supermatcher, svg, swiss, syco, tabix, table, tabular (highlighted in blue), tagseq, and tandem. To the right of the dropdown, there is a text box that says 'Changing dataset but not modify its contents. Change the type of your dataset.' On the right side of the interface, there is a 'History' panel showing a list of datasets. The datasets are numbered 1 through 7, with their names and some icons (eye, pencil, x) next to them. The dataset names include 'samples.txt', 'input.matrix.wt 37 vs wt GS NO.DESeq2.DE results', 'input.matrix.wt 37 vs wt ph8 .DESeq2.DE results', 'input.matrix.wt GSNO vs wt ph8.DESeq2.DE results', 'matrix.counts.matrix', 'de results', and '7: Extract and cluster differentially expressed transcripts on data 2, data 3, and others'.

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools search tools

Get Data Send Data Collection Operations COMMON TOOLS Text Manipulation Filter and Sort Join, Subtract and Group Convert Formats Extract Features Fetch Sequences Statistics Graph/Display Data Fasta Fastq Manipulation COMMON NGS TOOLS NGS:Samtools NGS:Mapping NGS:Bedtools NGS:Picard Tools SEARCHING TOOLS Diamond

Attributes Convert Format **Datatype** Permissions

Change data type

New Type:

txt
supermatcher
svg
swiss
syco
tabix
table
tabular
tagseq
tandem

Changing dataset but not modify its contents. Change the type of your dataset.

History

7: Extract and cluster differentially expressed transcripts on data 2, data 3, and others

7: Extract and cluster differentially expressed transcripts on data 2, data 3, and others: extracted differentially expressed genes
a list of datasets

6: de results
a list of 3 datasets

5: matrix.counts.matrix

4: input.matrix.wt GSNO vs wt ph8.DESeq2.DE results

3: input.matrix.wt 37 vs wt ph8 .DESeq2.DE results

2: input.matrix.wt 37 vs wt GS NO.DESeq2.DE results

1: samples.txt



1 Visualize this data

uc002zsw.2_cds_0_0_chr22_2

uc003bhh.3_cds_0_0_chr22_4



Trackster

Fast, interactive visualization for large, NGS/HTS datasets using only a web browser.



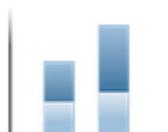
Bar Horizontal (NVD3)

Renders a horizontal bar diagram using NVD3 hosted at <http://www.nvd3.org>.



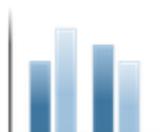
Bar Horizontal Stacked (NVD3)

Renders a stacked horizontal bar diagram using NVD3 hosted at <http://www.nvd3.org>.



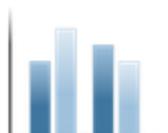
Bar Stacked (NVD3)

Renders a scatter plot using NVD3 hosted at <http://www.nvd3.org>.



Bar diagram (NVD3)

Renders a regular bar diagram using NVD3 hosted at <http://www.nvd3.org>.



Bar diagram (jqPlot)

Renders a bar diagram using jqPlot hosted at <http://www.jqplot.com>.



Box plot (jqPlot)

Processes tabular data using R and renders a box plot using jqPlot hosted at

7: Select first on data 6

5 lines

format: **tabular**, database: **hg38**



1	Tool Help	2
uc010gqp.3_cds_0_0_chr22_15690078_f	63	
uc011agd.3_cds_0_0_chr22_15528159_f	48	
uc062bek.1_cds_0_0_chr22_15690246_f	46	
uc003bhh.4_cds_0_0_chr22_46256561_r	30	
uc062bej.1_cds_1_0_chr22_15690426_f	26	

Tool help for Select first

What it does

This tool outputs specified number of lines from the **beginning** of a dataset

Example

Selecting 2 lines from this:

```
chr7 56632 56652 D17003_CTCF_R6
chr7 56736 56756 D17003_CTCF_R7
chr7 56761 56781 D17003_CTCF_R4
chr7 56772 56792 D17003_CTCF_R7
chr7 56775 56795 D17003_CTCF_R4
```

will produce:

```
chr7 56632 56652 D17003_CTCF_R6
chr7 56736 56756 D17003_CTCF_R7
```

HISTORY

History panel

Both inputs and outputs

The screenshot shows the Galaxy / ABiMS interface. The top navigation bar includes 'Galaxy / ABiMS', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', 'User', and 'Using 0%'. The left sidebar contains a 'Tools' section with a search bar and various tool categories like 'Get Data', 'Send Data', 'Collection Operations', 'COMMON TOOLS', 'COMMON NGS TOOLS', and 'SEARCHING TOOLS'. The main content area features a 'Welcome to galaxy3.sb-roscoff.fr' message, an 'Information' box with contact details, the 'ABiMS 4' logo, and a description of the Galaxy platform. The right sidebar is the 'History' panel, which lists datasets with their IDs, names, and actions. A red box highlights the top 8 datasets:

ID	Dataset Name	Actions
48	group2_count2.txt	View, Edit, Delete
47	group2_count1.txt	View, Edit, Delete
46	group1_count2.txt	View, Edit, Delete
45	group1_count1.txt	View, Edit, Delete
44	SARTools edgeR R objects (.RData)	View, Edit, Delete
43	SARTools edgeR R log	View, Edit, Delete
42	SARTools edgeR figures	View, Edit, Delete
41	SARTools edgeR tables	View, Edit, Delete
40	SARTools edgeR	View, Edit, Delete

History panel renaming and annotation

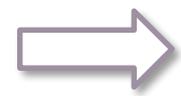
History ↻ ⚙ □

search datasets ✕

Unnamed History
28 shown, 14 deleted
1.59 MB ✓ 🗑 🗨

Click to rename history

- 48: group2_count2.txt 👁 ✎ ✕
- 47: group2_count1.txt 👁 ✎ ✕
- 46: group1_count2.txt 👁 ✎ ✕
- 45: group1_count1.txt 👁 ✎ ✕
- 44: SARTools edgeR R objects (.RData) 👁 ✎ ✕
- 43: SARTools edgeR R log 👁 ✎ ✕
- 42: SARTools edgeR figures 👁 ✎ ✕
- 41: SARTools edgeR tables 👁 ✎ ✕
- 40: SARTools edgeR 👁 ✎ ✕

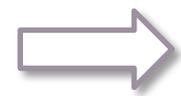


History ↻ ⚙ □

search datasets ✕

eba 2016 sartools
28 shown, 14 deleted
1.59 MB ✓ 🗑 🗨

- 48: group2_count2.txt 👁 ✎ ✕
- 47: group2_count1.txt 👁 ✎ ✕
- 46: group1_count2.txt 👁 ✎ ✕
- 45: group1_count1.txt 👁 ✎ ✕
- 44: SARTools edgeR R objects (.RData) 👁 ✎ ✕
- 43: SARTools edgeR R log 👁 ✎ ✕
- 42: SARTools edgeR figures 👁 ✎ ✕
- 41: SARTools edgeR tables 👁 ✎ ✕
- 40: SARTools edgeR 👁 ✎ ✕



History ↻ ⚙ □

search datasets ✕

eba 2016 sartools
28 shown, 14 deleted
1.59 MB ✓ 🗑 🗨

Tags:

Annotation:
 bla bla bla

- 48: group2_count2.txt 👁 ✎ ✕
- 47: group2_count1.txt 👁 ✎ ✕
- 46: group1_count2.txt 👁 ✎ ✕
- 45: group1_count1.txt 👁 ✎ ✕
- 44: SARTools edgeR R objects (.RData) 👁 ✎ ✕
- 43: SARTools edgeR R log 👁 ✎ ✕

History panel

History menu: Create new, Rename, Delete, **Delete Permanently**

The screenshot displays the Galaxy / ABiMS web interface. The top navigation bar includes 'Galaxy / ABiMS', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. A 'Using 0%' indicator is visible in the top right. The left sidebar contains a 'Tools' section with a search bar and various tool categories like 'Get Data', 'Send Data', 'Collection Operations', 'COMMON TOOLS', 'COMMON NGS TOOLS', and 'SEARCHING TOOLS'. The main content area features a green welcome message for 'galaxy3.sb-roscoff.fr', an information box with contact details, and the ABiMS logo with the tagline 'Analyses and Bioinformatics for Marine Science'. Below the logo are links for 'Changelog' and 'Tutorials', followed by a paragraph of text describing the Galaxy platform and its affiliations. On the right, the 'History' panel is open, showing a menu with categories: 'HISTORY LISTS' (Saved Histories, Histories Shared with Me), 'HISTORY ACTIONS' (Create New, Copy History, Share or Publish, Show Structure, Extract Workflow, Delete, Delete Permanently), 'DATASET ACTIONS' (Copy Datasets, Dataset Security, Resume Paused Jobs, Collapse Expanded Datasets, Unhide Hidden Datasets, Delete Hidden Datasets, Purge Deleted Datasets), 'DOWNLOADS' (Export Tool Citations, Export History to File), and 'OTHER ACTIONS' (Import from File). A red box highlights the gear icon in the History panel header.

History panel

Saved histories

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools search tools

Get Data Send Data Collection Operations COMMON TOOLS Text Manipulation Filter and Sort Join, Subtract and Group Convert Formats Extract Features Fetch Sequences Statistics Graph/Display Data Fasta Fastq Manipulation COMMON NGS TOOLS NGS:Samtools NGS:Mapping NGS:Bedtools NGS:Picard Tools SEARCHING TOOLS Diamond

search history names and tags

Advanced Search

Name	Datasets	Tags	Sharing	Size on Disk	Created	Last
eba2016 deseq2	8	0 Tags		829.2 KB	Nov 14, 2016	Nov 1
Copy of TP_ITMO2016 shared by xi.liu@sb- roscoff.fr (active items only)	32 2	0 Tags		28.5 MB	Nov 10, 2016	Nov 1
eba 2016 macs2	3 2	0 Tags		602.8 KB	Nov 09, 2016	Nov C
eba 2016 sickle	10	0 Tags		4.2 MB	Oct 04, 2016	Nov C

Page: 1 2 | Show All

For 0 selected histories: Rename Delete Delete Permanently Undele

History

HISTORY LISTS

- Saved Histories
- Histories Shared with Me

HISTORY ACTIONS

- Create New
- Copy History
- Share or Publish
- Show Structure
- Extract Workflow
- Delete
- Delete Permanently

DATASET ACTIONS

- Copy Datasets
- Dataset Security
- Resume Paused Jobs
- Collapse Expanded Datasets
- Unhide Hidden Datasets
- Delete Hidden Datasets
- Purge Deleted Datasets

DOWNLOADS

- Export Tool Citations
- Export History to File

OTHER ACTIONS

- Import from File

galaxy3.sb-roscoff.fr/history/list

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

History panel

Saved histories: Switch histories

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools search tools

Get Data Send Data Collection Operations COMMON TOOLS Text Manipulation Filter and Sort Join, Subtract and Group Convert Formats Extract Features Fetch Sequences Statistics Graph/Display Data Fasta Fastq Manipulation COMMON NGS TOOLS NGS:Samtools NGS:Mapping NGS:Bedtools NGS:Picard Tools SEARCHING TOOLS Diamond

search history names and tags

Advanced Search

<input type="checkbox"/>	Name	Datasets	Tags	Sharing	Size on Disk	Created	Last
<input type="checkbox"/>	eba2016 deseq2	8	0 Tags		829.2 KB	Nov 14, 2016	Nov 14, 2016
<input type="checkbox"/>	Copy of TP_ITMO2016 shared by xi.liu@sb-roscoff.fr (active items only)	32 2	0 Tags		28.5 MB	Nov 10, 2016	Nov 10, 2016
<input type="checkbox"/>	eba 2016 macs2	3 2	0 Tags		602.8 KB	Nov 09, 2016	Nov 09, 2016
<input type="checkbox"/>	eba 2016 sickle	10	0 Tags		4.2 MB	Oct 04, 2016	Nov 04, 2016

Page: 1 2 | Show All

For 0 selected histories: Rename Delete Delete Permanently Undelete

History

HISTORY LISTS

- Saved Histories
- Histories Shared with Me

HISTORY ACTIONS

- Create New
- Copy History
- Share or Publish
- Show Structure
- Extract Workflow
- Delete
- Delete Permanently

DATASET ACTIONS

- Copy Datasets
- Dataset Security
- Resume Paused Jobs
- Collapse Expanded Datasets
- Unhide Hidden Datasets
- Delete Hidden Datasets
- Purge Deleted Datasets

DOWNLOADS

- Export Tool Citations
- Export History to File

OTHER ACTIONS

- Import from File

galaxy3.sb-roscoff.fr/history/list

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

History panel

History menu: Create new, Rename, Delete, **Delete Permanently**

The screenshot displays the Galaxy/ABiMS web interface. The top navigation bar includes 'Galaxy / ABiMS', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. A 'Using 0%' indicator is visible in the top right. The left sidebar contains a 'Tools' section with a search bar and various tool categories like 'Get Data', 'Send Data', 'Collection Operations', 'COMMON TOOLS', 'COMMON NGS TOOLS', and 'SEARCHING TOOLS'. The main content area features a green welcome banner for 'galaxy3.sb-roscoff.fr', an information box with contact details, and the ABiMS logo with the tagline 'Analyses and Bioinformatics for Marine Science'. Below the logo are links for 'Changelog' and 'Tutorials', followed by a paragraph of text about the Galaxy platform. On the right, the 'History' panel is open, showing a list of actions. The 'HISTORY ACTIONS' section is highlighted with a red box, and the 'Create New' option is also highlighted with a red box. Other actions include 'Copy History', 'Share or Publish', 'Show Structure', 'Extract Workflow', 'Delete', 'Delete Permanently', 'DATASET ACTIONS', 'Copy Datasets', 'Dataset Security', 'Resume Paused Jobs', 'Collapse Expanded Datasets', 'Unhide Hidden Datasets', 'Delete Hidden Datasets', 'Purge Deleted Datasets', 'DOWNLOADS', 'Export Tool Citations', 'Export History to File', and 'OTHER ACTIONS' with 'Import from File'.

History panel

The screenshot displays the Galaxy / ABiMS web interface. The top navigation bar includes 'Galaxy / ABiMS', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', 'User', and a 'Using 0%' indicator. The left sidebar contains a 'Tools' section with a search bar and various tool categories: 'Get Data', 'Send Data', 'Collection Operations', 'COMMON TOOLS' (Text Manipulation, Filter and Sort, Join, Subtract and Group, Convert Formats, Extract Features, Fetch Sequences, Statistics, Graph/Display Data, Fasta Fastq Manipulation), and 'COMMON NGS TOOLS' (NGS:Samtools, NGS:Mapping, NGS:Bedtools, NGS:Picard Tools). The main content area features a green welcome message for 'galaxy3.sb-roscoff.fr', an information box with contact details, the ABiMS logo (Analyses and Bioinformatics for Marine Science), and links to 'Changelog' and 'Tutorials'. A paragraph of text describes Galaxy as an open, web-based platform for data-intensive biomedical research, supported by various institutions. The right sidebar is the 'History' panel, which lists datasets with search, view, edit, and delete icons. The top of the history panel has a search bar and a red box highlighting a refresh, settings, and close icon. The history list includes: 'eba 2016 sartools' (1.59 MB), '48: group2_count2.txt' (marked with a red X), '47: group2_count1.txt', '46: group1_count2.txt', '45: group1_count1.txt', '44: SARTools edgeR R objects (.RData)', '43: SARTools edgeR R log', '42: SARTools edgeR figures', '41: SARTools edgeR tables', and '40: SARTools edgeR'.

History panel

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Done search histories search all datasets Create new

Current History

Switch to

eba 2016 sartools
28 shown, 14 [deleted](#)
1.59 MB

search datasets

Drag datasets here to copy them to the current history

- 48: [group2_count2.txt](#)
- 47: [group2_count1.txt](#)
- 46: [group1_count2.txt](#)
- 45: [group1_count1.txt](#)
- 44: [SARTools edgeR R objects \(.RData\)](#)
- 43: [SARTools edgeR R log](#)
- 42: [SARTools edgeR figures](#)
- 41: [SARTools edgeR tables](#)
- 40: [SARTools edgeR report](#)

Switch to

Trinity example
3 shown, 3 [deleted](#)
40.3 KB

search datasets

- 5: [Trinity on data 3 and data 4: Assembled Transcripts](#)
- 4: [reads.left.fq](#)
- 3: [reads.right.fq](#)

Switch to

trinity_contig_exn50_statistic
12 shown, 15 [deleted](#)
47.01 KB

search datasets

- 14: [Build expression matrix on data 7 and data 6: matrix of UpperQuartile-normalized expression values](#)
- 13: [Build expression matrix on data 7 and data 6: matrix of TPM expression values \(not cross-sample normalized\)](#)
- 12: [Build expression matrix on data 7 and data 6: estimated RNA-Seq fragment counts \(raw counts\)](#)
- 9: [Build expression matrix on data 7 and data 6: matrix of TPM expression values \(not cross-sample normalized\)](#)
- 8: [Build expression matrix on data 7 and data 6: estimated](#)

Switch to

eba 2016 tr
16 shown
21.92 KB

search da

- 16: [Extra](#)
- 15: [Extract](#)
- 14: [Extract](#)
- 13: [Extra](#)
- 12: [Extra](#)

History panel

The screenshot shows the Galaxy / ABiMS interface. At the top, there is a navigation bar with 'Galaxy / ABiMS' and various menu items: 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. A 'Using 0%' indicator is on the right. Below the navigation bar, there are search bars for 'search histories' and 'search all datasets'. The main area is titled 'Current History' and contains several history items, each with a 'Switch to' button. The first item is 'eba 2016 sartools' (1.59 MB, 28 shown, 14 deleted). The second item is 'Trinity example' (40.3 KB, 3 shown, 3 deleted), which is currently selected. The third item is 'trinity_contig_exn50_statistic' (47.01 KB, 12 shown, 15 deleted). The fourth item is 'eba 2016 tr...' (21.92 KB, 16 shown). Each history item has a search bar and a 'Switch to' button. The 'Trinity example' item is highlighted with a red box around its 'Switch to' button. The history items are listed in a table-like format with columns for item name, size, and status. The items are: 48: group2_count2.txt (red background), 47: group2_count1.txt (green background), 46: group1_count2.txt (green background), 45: group1_count1.txt (green background), 44: SARTools edgeR R objects (.RData) (green background), 43: SARTools edgeR R log (green background), 42: SARTools edgeR figures (green background), 41: SARTools edgeR tables (green background), 40: SARTools edgeR report (green background), 5: Trinity on data 3 and data 4: Assembled Transcripts (green background), 4: reads.left.fg (green background), 3: reads.right.fg (green background), 14: Build expression matrix on data 7 and data 6: matrix of UpperQuartile-normalized expression values (green background), 13: Build expression matrix on data 7 and data 6: matrix of TPM expression values (not cross-sample normalized) (green background), 12: Build expression matrix on data 7 and data 6: estimated RNA-Seq fragment counts (raw counts) (green background), 9: Build expression matrix on data 7 and data 6: matrix of TPM expression values (not cross-sample normalized) (green background), 8: Build expression matrix on data 7 and data 6: estimated (green background), 16: Extra differentially transcripts of RData file (red background), 15: Extract expressed to data 3, and depleted cat a list of datase (green background), 14: Extract expressed to data 3, and a list of datase (green background), 13: Extra differentially transcripts of (red background), 12: Extra differentially (red background).

History panel

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Done search histories search all datasets Create new

Current History

Switch to

eba 2016 sartools
28 shown, 14 deleted
1.59 MB

search datasets

Drag datasets here to copy them to the current history

- 48: group2_count2.txt
- 47: group2_count1.txt
- 46: group1_count2.txt
- 45: group1_count1.txt
- 44: SARTools edgeR R objects (.RData)
- 43: SARTools edgeR R log
- 42: SARTools edgeR figures
- 41: SARTools edgeR tables
- 40: SARTools edgeR report

Switch to

Trinity example
3 shown, 3 deleted
40.3 KB

search datasets

- 5: Trinity on data 3 and data 4: Assembled Transcripts
- 4: reads.left.fg
- 3: reads.right.fg

Switch to

trinity_contig_exn50_statistic
12 shown, 15 deleted
47.01 KB

search datasets

- 14: Build expression matrix on data 7 and data 6: matrix of UpperQuartile-normalized expression values
- 13: Build expression matrix on data 7 and data 6: matrix of TPM expression values (not cross-sample normalized)
- 12: Build expression matrix on data 7 and data 6: estimated RNA-Seq fragment counts (raw counts)
- 9: Build expression matrix on data 7 and data 6: matrix of TPM expression values (not cross-sample normalized)
- 8: Build expression matrix on data 7 and data 6: estimated

Switch to

eba 2016 tr
16 shown
21.92 KB

search datasets

- 16: Extra differentially transcripts of RData file
- 15: Extract expressed to data 3, and depleted cat a list of dataset
- 14: Extract expressed to data 3, and a list of dataset
- 13: Extra differentially transcripts of
- 12: Extra differentially

History panel

The screenshot displays the Galaxy / ABiMS interface. At the top, there is a navigation bar with the following items: Galaxy / ABiMS, Analyze Data, Workflow, Shared Data, Visualization, Admin, Help, User, and a grid icon. On the right side of the navigation bar, it says "Using 0%" and "Create new". Below the navigation bar, there is a search bar with two input fields: "search histories" and "search all datasets". A red box highlights these two search fields. Below the search bar, there are three panels, each with a "Switch to" dropdown menu. The first panel is titled "Current History" and shows a list of datasets. The second panel is titled "Trinity example" and shows a list of datasets. The third panel is titled "trinity_contig_exn50_statistic" and shows a list of datasets. The fourth panel is partially visible on the right and is titled "eba 2016 tr". Each panel has a search bar and a list of datasets with icons for viewing, editing, and deleting. The datasets in the "Current History" panel include: 48: group2_count2.txt, 47: group2_count1.txt, 46: group1_count2.txt, 45: group1_count1.txt, 44: SARTools edgeR R objects (.RData), 43: SARTools edgeR R log, 42: SARTools edgeR figures, 41: SARTools edgeR tables, and 40: SARTools edgeR report. The datasets in the "Trinity example" panel include: 5: Trinity on data 3 and data 4: Assembled Transcripts, 4: reads.left.fq, and 3: reads.right.fq. The datasets in the "trinity_contig_exn50_statistic" panel include: 14: Build expression matrix on data 7 and data 6: matrix of UpperQuartile-normalized expression values, 13: Build expression matrix on data 7 and data 6: matrix of TPM expression values (not cross-sample normalized), 12: Build expression matrix on data 7 and data 6: estimated RNA-Seq fragment counts (raw counts), 9: Build expression matrix on data 7 and data 6: matrix of TPM expression values (not cross-sample normalized), and 8: Build expression matrix on data 7 and data 6: estimated.

History panel

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Done search histories search all datasets Create new

Current History

Switch to

eba 2016 sartools
28 shown, 14 [deleted](#)
1.59 MB

search datasets

Drag datasets here to copy them to the current history

- 48: [group2_count2.txt](#)
- 47: [group2_count1.txt](#)
- 46: [group1_count2.txt](#)
- 45: [group1_count1.txt](#)
- 44: [SARTools edgeR R objects \(.RData\)](#)
- 43: [SARTools edgeR R log](#)
- 42: [SARTools edgeR figures](#)
- 41: [SARTools edgeR tables](#)
- 40: [SARTools edgeR report](#)

Trinity example
3 shown, 3 [deleted](#)
40.3 KB

search datasets

- 5: [Trinity on data 3 and data 4: Assembled Transcripts](#)
- 4: [reads.left.fg](#)
- 3: [reads.right.fg](#)

trinity_contig_exn50_statistic
12 shown, 15 [deleted](#)
47.01 KB

search datasets

- 14: [Build expression matrix on data 7 and data 6: matrix of UpperQuartile-normalized expression values](#)
- 13: [Build expression matrix on data 7 and data 6: matrix of TPM expression values \(not cross-sample normalized\)](#)
- 12: [Build expression matrix on data 7 and data 6: estimated RNA-Seq fragment counts \(raw counts\)](#)
- 9: [Build expression matrix on data 7 and data 6: matrix of TPM expression values \(not cross-sample normalized\)](#)
- 8: [Build expression matrix on data 7 and data 6: estimated](#)

eba 2016 tr
16 shown
21.92 KB

search da

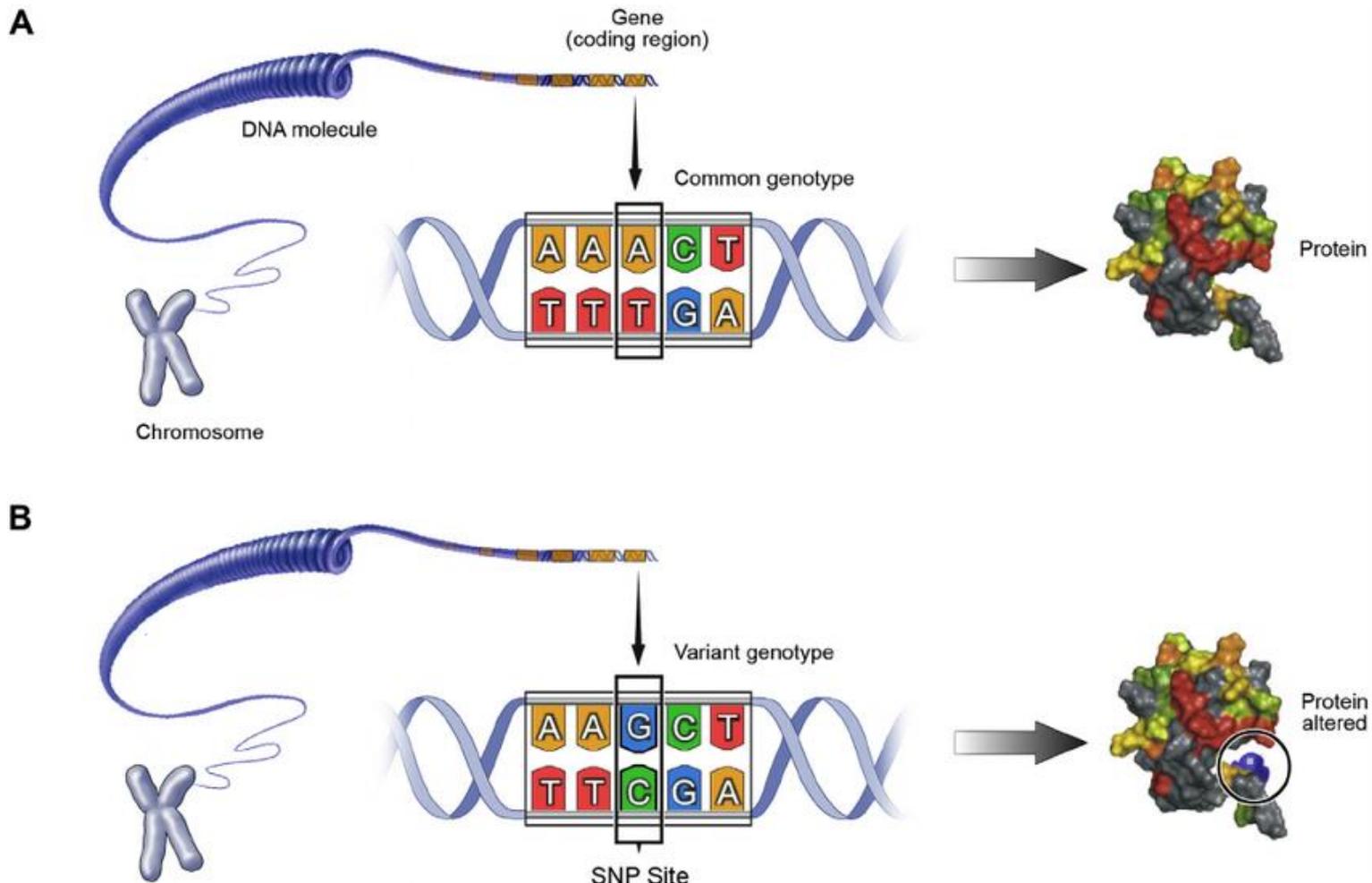
- 16: [Extra differentially transcripts of RData file](#)
- 15: [Extract expressed to data 3, and depleted cat a list of datase](#)
- 14: [Extract expressed to data 3, and a list of datase](#)
- 13: [Extra differentially transcripts of](#)
- 12: [Extra differentially](#)

Hands-on
TOOLS





Which coding exon has the highest number of single nucleotide polymorphisms on chromosome 22?





Which coding exon has the highest number of single nucleotide polymorphisms on chromosome 22?

1. Get the data in a new history
2. Join exons with SNPs
3. Count the number of SNPs per exon
4. Sort exons by SNP count
5. Select top five
6. Recover exon info
7. Display data in UCSC genome browser



Which coding exon has the highest number of single nucleotide polymorphisms on chromosome 22?

1. Get the data in a new history
2. Join exons with SNPs *Operate on Genomics Intervals -> Join*
3. Count the number of SNPs per exon
4. Sort exons by SNP count
5. Select top five
6. Recover exon info
7. Display data in UCSC genome browser



Which coding exon has the highest number of single nucleotide polymorphisms on chromosome 22?

1. Get the data in a new history
2. Join exons with SNPs *Operate on Genomics Intervals -> Join*
3. Count the number of SNPs per exon *Join, Subtract, and Group -> Group*
4. Sort exons by SNP count
5. Select top five
6. Recover exon info
7. Display data in UCSC genome browser



Which coding exon has the highest number of single nucleotide polymorphisms on chromosome 22?

1. Get the data in a new history
2. Join exons with SNPs *Operate on Genomics Intervals -> Join*
3. Count the number of SNPs per exon *Join, Subtract, and Group -> Group*
4. Sort exons by SNP count *Filter and Sort -> Sort*
5. Select top five
6. Recover exon info
7. Display data in UCSC genome browser



Which coding exon has the highest number of single nucleotide polymorphisms on chromosome 22?

1. Get the data in a new history
2. Join exons with SNPs *Operate on Genomics Intervals -> Join*
3. Count the number of SNPs per exon *Join, Subtract, and Group -> Group*
4. Sort exons by SNP count *Filter and Sort -> Sort*
5. Select top five *Text Manipulation -> Select First*
6. Recover exon info
7. Display data in UCSC genome browser



Which coding exon has the highest number of single nucleotide polymorphisms on chromosome 22?

1. Get the data in a new history
2. Join exons with SNPs *Operate on Genomics Intervals -> Join*
3. Count the number of SNPs per exon *Join, Subtract, and Group -> Group*
4. Sort exons by SNP count *Filter and Sort -> Sort*
5. Select top five *Text Manipulation -> Select First*
6. Recover exon info *Join, Subtract and Group -> Compare two Datasets*
7. Display data in UCSC genome browser

Cleanup

DATASET



Dataset

Delete a dataset

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools

search tools

Get Data
Send Data
Collection Operations

COMMON TOOLS
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Statistics
Graph/Display Data
Fasta Fastq Manipulation

COMMON NGS TOOLS
NGS:Samtools
NGS:Mapping
NGS:Bedtools
NGS:Picard Tools

SEARCHING TOOLS
Diamond

Welcome to galaxy3.sb-roscoff.fr

Information
For any question or request for tools or account, send an email at support.abims@sb-roscoff.fr

ABiMS
Analyses and Bioinformatics for Marine Science

Station Biologique Roscoff

Changelog
Tutorials

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

search datasets

eba 2016 sartools
41 shown, 1 deleted
1.59 MB

62: SARTools DESeq2 R objects (.RData) **Delete**

61: SARTools DESeq2 R log

60: SARTools DESeq2 figures

59: SARTools DESeq2 tables

58: SARTools DESeq2 report

57: SARTools edgeR R objects (.RData)

56: SARTools edgeR R log

55: SARTools edgeR figures

javascript:void(0);

Dataset



The dataset isn't really deleted. It's in the Trash

The screenshot shows the Galaxy / ABiMS web interface. The top navigation bar includes 'Galaxy / ABiMS', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. A 'Using 0%' indicator is in the top right. The left sidebar contains a 'tools' section with a search bar and various tool categories like 'Get Data', 'Send Data', 'Collection Operations', 'COMMON TOOLS', 'COMMON NGS TOOLS', and 'SEARCHING TOOLS'. The main content area features a green welcome banner for 'galaxy3.sb-roscoff.fr', an information box with contact details, the ABiMS logo (Analyses and Bioinformatics for Marine Science), and a list of links for 'Changelog' and 'Tutorials'. The right sidebar shows a 'History' panel with a search bar and a list of datasets. The dataset 'eba 2016 sartools' is highlighted with a red box, showing '42 shown, hide deleted' and '1.59 MB'. Below it, a yellow warning box with a red border states: 'This dataset has been deleted. Undelete it. Permanently remove it from disk'. The history list includes items like '62: SARTools DESeq2 R objects (.RData)', '61: SARTools DESeq2 R log', '60: SARTools DESeq2 figures', '59: SARTools DESeq2 tables', '58: SARTools DESeq2 report', '57: SARTools edgeR R objects (.RData)', and '56: SARTools edgeR'. The bottom of the interface shows a JavaScript console with 'javascript:void(0);'.

Dataset



“Empty Trash” : to free up disk space

The screenshot shows the Galaxy / ABiMS web interface. The top navigation bar includes 'Galaxy / ABiMS', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', 'User', and 'Using 0%'. The left sidebar contains a 'Tools' section with a search bar and various tool categories like 'Get Data', 'Send Data', 'Collection Operations', 'COMMON TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Convert Formats', 'Extract Features', 'Fetch Sequences', 'Statistics', 'Graph/Display Data', 'Fasta Fastq Manipulation', 'COMMON NGS TOOLS', 'NGS:Samtools', 'NGS:Mapping', 'NGS:Bedtools', 'NGS:Picard Tools', 'SEARCHING TOOLS', and 'Diamond'. The main content area features a green welcome banner for 'galaxy3.sb-roscoff.fr', an information box with contact details, the ABiMS logo (Analyses and Bioinformatics for Marine Science), and links to 'Changelog' and 'Tutorials'. A paragraph of text describes Galaxy as an open, web-based platform for data intensive biomedical research, supported by various institutions. On the right, the 'History' panel is open, showing a list of actions: 'HISTORY LISTS', 'Saved Histories', 'Histories Shared with Me', 'HISTORY ACTIONS', 'Create New', 'Copy History', 'Share or Publish', 'Show Structure', 'Extract Workflow', 'Delete', 'Delete Permanently', 'DATASET ACTIONS', 'Copy Datasets', 'Dataset Security', 'Resume Paused Jobs', 'Collapse Expanded Datasets', 'Unhide Hidden Datasets', 'Delete Hidden Datasets', 'Purge Deleted Datasets' (highlighted), 'DOWNLOADS', 'Export Tool Citations', 'Export History to File', and 'OTHER ACTIONS', 'Import from File'. The browser address bar at the bottom shows 'galaxy3.sb-roscoff.fr/history/purge_deleted_datasets'.

Datatypes

DATASET

Common text formats:

- ***txt***: plain text ('.txt')
- ***tabular***: tab delimited ('.tab', '.txt', etc.)

	wt_37_2	wt_37_3	wt_37_1
TR24 c0_g1_i1	90.00	67.00	85.00
TR2779 c0_g1_i1	186.00	137.00	217.00
TR127 c1_g1_i1	9.00	23.00	16.00

- ***csv***: comma-separated values ('.csv')

```
Year,Make,Model
1997,Ford,E350
2000,Mercury,Cougar
```

- ***html***: standard language for web pages

```
<!DOCTYPE html>
<html>
  <head>
    <title>This is a title</title>
  </head>
  <body>
    <p>Hello world!</p>
  </body>
</html>
```

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools

search tools

Get Data
Send Data
Collection Operations

COMMON TOOLS

Text Manipulation

Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Statistics
Graph/Display Data
Fasta Fastq Manipulation

COMMON NGS TOOLS

NGS:Samtools
NGS:Mapping
NGS:Bedtools
NGS:Picard Tools

SEARCHING TOOLS

Diamond

Information
For any support, send an email at

Text Manipulation

- Add column to an existing dataset
- Concatenate datasets tail-to-head
- Cut columns from a table
- Merge Columns together
- Convert delimiters to TAB
- Create single interval as a new dataset
- Change Case of selected columns
- Paste two files side by side
- Remove beginning of a file
- Select random lines from a file

History

search datasets

eba 2016 sartools
42 shown
1.59 MB

62: SARTools DESeq2 R objects (.RData)

61: SARTools DESeq2 R log

60: SARTools DESeq2 figures

59: SARTools DESeq2 tables

58: SARTools DESeq2 report

57: SARTools edgeR R objects (.RData)

56: SARTools edgeR R log

55: SARTools edgeR figures

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools

search tools

Get Data

Send Data

Collection Operations

COMMON TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

Extract Features

Fetch Sequences

Statistics

Graph/Display Data

Fasta Fastq Manipulation

COMMON NGS TOOLS

NGS:Samtools

NGS:Mapping

NGS:Bedtools

NGS:Picard Tools

SEARCHING TOOLS

Diamond

✓ Welcome to galaxy3.sb-roscoff.fr

Information

For any support, send an email at

Filter and Sort

Filter data on any column using simple expressions

Sort data in ascending or descending order

Select lines that match an expression

GFF

Extract features from GFF data

Filter GFF data by attribute using simple expressions

Filter GFF data by feature count using simple expressions

Filter GTF data by attribute

Changelog

Tutorials

Galaxy is an open source platform for collaborative biomedical research. The Galaxy team is a multidisciplinary team in Computer Science, Biology and Mathematics and The Galaxy Project is supported in part by NHGRI, the National Science Foundation, The Institute for CyberScience at Penn State, and Emory University.

History

search datasets

eba 2016 sartools

42 shown

1.59 MB

- 62: SARTools DESeq2 R objects (.RData)**
- 61: SARTools DESeq2 R log**
- 60: SARTools DESeq2 figures**
- 59: SARTools DESeq2 tables**
- 58: SARTools DESeq2 report**
- 57: SARTools edgeR R objects (.RData)**
- 56: SARTools edgeR R log**
- 55: SARTools edgeR figures**

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools

search tools

Get Data

Send Data

Collection Operations

COMMON TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

Extract Features

Fetch Sequences

Statistics

Graph/Display Data

Fasta Fastq Manipulation

COMMON NGS TOOLS

NGS:Samtools

NGS:Mapping

NGS:Bedtools

NGS:Picard Tools

SEARCHING TOOLS

Diamond

✓ Welcome to galaxy3.sb-roscoff.fr

Information

For any support, send an email at [support@roscoff.fr](#)

Join, Subtract and Group

Join two Datasets side by side on a specified field

Compare two Datasets to find common or distinct rows

Group data by a column and perform aggregate operation on other columns.



Analyses and Bioinformatics for Marine Science

▶ Changelog

▶ Tutorials

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

search datasets

eba 2016 sartools
42 shown
1.59 MB

62: SARTools DESeq2 R objects (.RData)

61: SARTools DESeq2 R log

60: SARTools DESeq2 figures

59: SARTools DESeq2 tables

58: SARTools DESeq2 report

57: SARTools edgeR R objects (.RData)

56: SARTools edgeR R log

55: SARTools edgeR figures

Common binary formats:

- ***data***: generic binary format
- ***zip, tar***: archives
- ***pdf, png, jpg, bmp, tiff, gif***: images
- ***rdata***: statistical computing program R
- ***bam, wig, bigwig***: sequence alignment

Sequence file formats:

- ***fasta***: a single-line description with '>', followed by lines of sequence data ('.fasta', '.fas')

```
>sequence1
atgcgtttgcgtgcatgcgtttgcgtgcatgcgtttgcgtgcatgcgtttgcgtgc
atgcgtttgcgtgc
>sequence2
tttcgtgcgatatagtttcgtgcgatatagtttcgtgcgatatag
tggcgcggt
```

- ***fastq***: sequence + quality score ('.fastq', '.fq')

```
@SEQ_ID
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTTT
+
!''*(((((***+))%%%++) (%%%) .1***-+*'')) **55CCF>>>>>CCCCCCC65
@SEQ_ID2
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTTT
+
!''*(((((***+))%%%++) (%%%) .1***-+*'')) **55CCF>>>>>CCCCCCC65
```

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools

search tools

Get Data

Send Data

Collection Operations

COMMON TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

Extract Features

Fetch Sequences

Statistics

Graph/Display Data

Fasta Fastq Manipulation

COMMON NGS TOOLS

NGS:Samtools

NGS:Mapping

NGS:Bedtools

NGS:Picard Tools

SEARCHING TOOLS

Diamond

Welcome to galaxy3.sb-roscoff.fr

Information
 For any question or request for tools or account, send an email at support.abims@sb-roscoff.fr



Analyses and Bioinformatics for Marine Science

Changelog

Tu

Galaxy
Galaxy
Comput
in part
CyberS

data intensive biomedical research. The
and the [Biology](#) and [Mathematics](#) and
iversity. The [Galaxy Project](#) is supported
[the Life Sciences](#), [The Institute for](#)
ersity.

History

search datasets

eba 2016 sartools
42 shown
1.59 MB

62: SARTools DESeq2 R objects (.RData)

61: SARTools DESeq2 R log

60: SARTools DESeq2 figures

59: SARTools DESeq2 tables

58: SARTools DESeq2 report

57: SARTools edgeR R objects (.RData)

56: SARTools edgeR R log

55: SARTools edgeR figures

Fasta Fastq Manipulation

- [faSplit](#) Split a FASTA file
- [FastQC](#) Read Quality reports
- [Filter sequences by length](#)
- [FASTQ Groomer](#) convert between various FASTQ quality formats
- [Filter sequences by ID](#) from a tabular file

Sequence file formats:

- ***gff3*, *bed*, *genbank***: sequence + annotations

bed

```
track name=pairedReads description="Clone Paired Reads" useScore=1
chr22 1000 5000 cloneA 960 + 1000 5000 0 2 567,488, 0,3512
chr22 2000 6000 cloneB 900 - 2000 6000 0 2 433,399, 0,3601
```

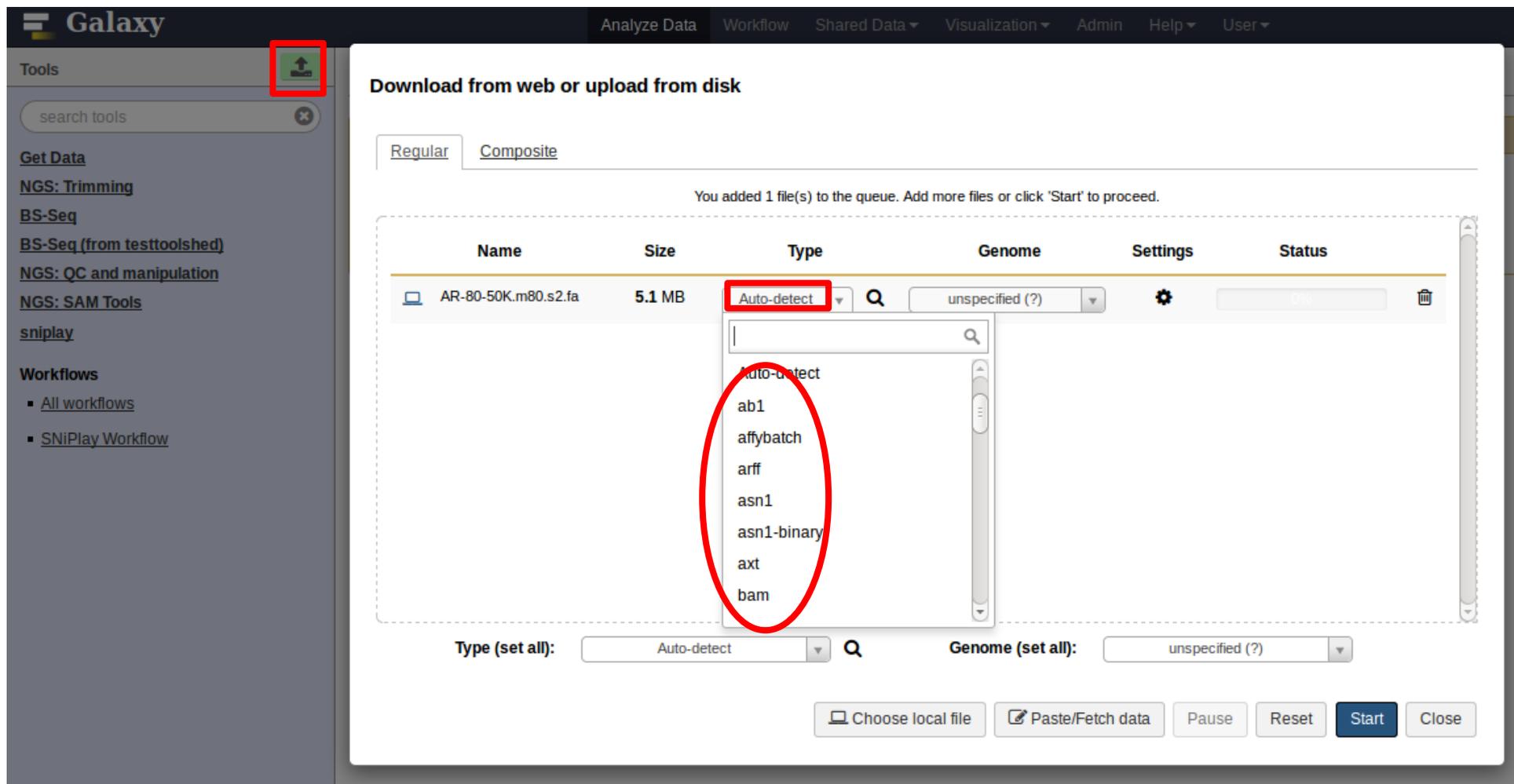
gff3

```
##gff-version 3
ctg123 . exon 1300 1500 . + . ID=exon00001
ctg123 . exon 1050 1500 . + . ID=exon00002
ctg123 . exon 3000 3902 . + . ID=exon00003
##FASTA
>ctg123
cttctgggcgtacccgattctcggagaacttgccgcaccattccgccttg
tgttcattgctgcctgcatgttcattgtctacctcggctacgtgtggcta
...
```

<https://genome.ucsc.edu/FAQ/FAQformat.html#format3>

Dataset - Datatypes

- Every Galaxy dataset is associated with a datatype.
- Datatype can be detected or user specified.



Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help User

Tools

search tools

Get Data

NGS: Trimming

BS-Seq

BS-Seq (from testtoolshed)

NGS: QC and manipulation

NGS: SAM Tools

sniplay

Workflows

- All workflows
- SNiPlay Workflow

Download from web or upload from disk

Regular Composite

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

Name	Size	Type	Genome	Settings	Status
AR-80-50K.m80.s2.fa	5.1 MB	Auto-detect	unspecified (?)		

Type (set all): Auto-detect Genome (set all): unspecified (?)

Choose local file Paste/Fetch data Pause Reset Start Close

Dataset - Datatypes

- Every Galaxy dataset is associated with a datatype.
- Datatype can be detected or user specified.

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools search tools

Get Data
Send Data
Collection Operations

COMMON TOOLS
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Statistics
Graph/Display Data
Fasta Fastq Manipulation

COMMON NGS TOOLS
NGS:Samtools
NGS:Mapping
NGS:Bedtools
NGS:Picard Tools

SEARCHING TOOLS
Diamond

Attributes **Datatype** Permissions

Change data type

New Type:

txt
supermatcrier
svg
swiss
syco
tabix
table
tabular
tagseq
tandem

ing dataset but *not* modify its contents.
d the type of your dataset.

History

7: Extract and cluster differentially expressed transcripts on data 2, data 3, and others

7: Extract and cluster differentially expressed transcripts on data 2, data 3, and others: extracted differentially expressed genes
a list of datasets

6: de results
a list of 3 datasets

5: matrix.counts.matrix

4: input.matrix.wt GSNO vs wt ph8.DESeq2.DE results

3: input.matrix.wt 37 vs wt ph8.DESeq2.DE results

2: input.matrix.wt 37 vs wt GSNO.DESeq2.DE results

1: samples.txt

- Many tools will only accept input datasets with the appropriate datatype assigned.

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools

trinity

Trinity suite

1- ASSEMBLY:

[Trinity](#) de novo assembly of RNA-Seq data

[Trinity Statistics](#) Obtain basic stats for the number of genes and isoforms and contiguity of the assembly

[Generate gene to transcript map](#) for Trinity assembly

2- COUNTING:

[Align reads and estimate abundance](#) on a de novo assembly of RNA-Seq data

[Build expression matrix](#) for a de novo assembly of RNA-Seq data by Trinity

4- ANNOTATION:

[Filter low expression transcripts](#) from a Trinity assembly

[TransDecoder](#) Find coding

Trinity de novo assembly of RNA-Seq data (Galaxy Version 2.2.0.0) Options

Paired or Single-end data?

Paired

Left/Forward strand reads

No fasta or fastqsanger dataset available.

(--left)

Right/Reverse strand reads

No fasta or fastqsanger dataset available.

(--right)

Strand specific data

Yes No

Jaccard Clip options

Yes No

set if you expect high gene density with UTR overlap (--jaccard_clip)

Run in silico normalization of reads

Yes No

Defaults to max. read coverage of 50. (--normalize_reads)

History

search datasets

eba 2016 sartools
40 shown, 2 deleted
1.59 MB

61: SARTools DESeq2 R log

60: SARTools DESeq2 figures

59: SARTools DESeq2 tables

58: SARTools DESeq2 report

57: SARTools edgeR R objects (.RData)

56: SARTools edgeR R log

55: SARTools edgeR figures

54: SARTools edgeR tables

53: SARTools edgeR

- Many tools will only accept input datasets with the appropriate datatype assigned.

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools

trinity

Trinity suite

1- ASSEMBLY:

Trinity de novo assembly of RNA-Seq data

Trinity Statistics Obtain basic stats for the number of genes and isoforms and contiguity of the assembly

Generate gene to transcript map for Trinity assembly

2- COUNTING:

Align reads and estimate abundance on a de novo assembly of RNA-Seq data

Build expression matrix for a de novo assembly of RNA-Seq data by Trinity

4- ANNOTATION:

TransDecoder Find coding

Trinity de novo assembly of RNA-Seq data (Galaxy Version 2.2.0.0) Options

Paired or Single-end data?

Paired

Left/Forward strand reads

No fasta or fastqsanger dataset available.

(--left)

Right/Reverse strand reads

No fasta or fastqsanger dataset available.

(--right)

Strand specific data

Yes No

Jaccard Clip options

Yes No

History

search datasets

eba 2016 sartools
40 shown, 2 deleted
1.59 MB

61: SARTools DESeq2 R log

60: SARTools DESeq2 figures

59: SARTools DESeq2 tables

58: SARTools DESeq2 report

57: SARTools edgeR R objects (.RData)

56: SARTools edgeR R log

55: SARTools edgeR figures

53: SARTools edgeR

Defaults to max. read coverage of 50. (--normalize_reads)

<https://galaxyproject.github.io/training-material/topics/introduction/tutorials/galaxy-intro-ngs-data-managment/tutorial.html#what-is-fastq>

Hands-on

TOOLS (sequence files manipulation)





Part 1:

You have sequencing data from your favorite species. You want to check the quality of your sequences and to map on the reference genome.

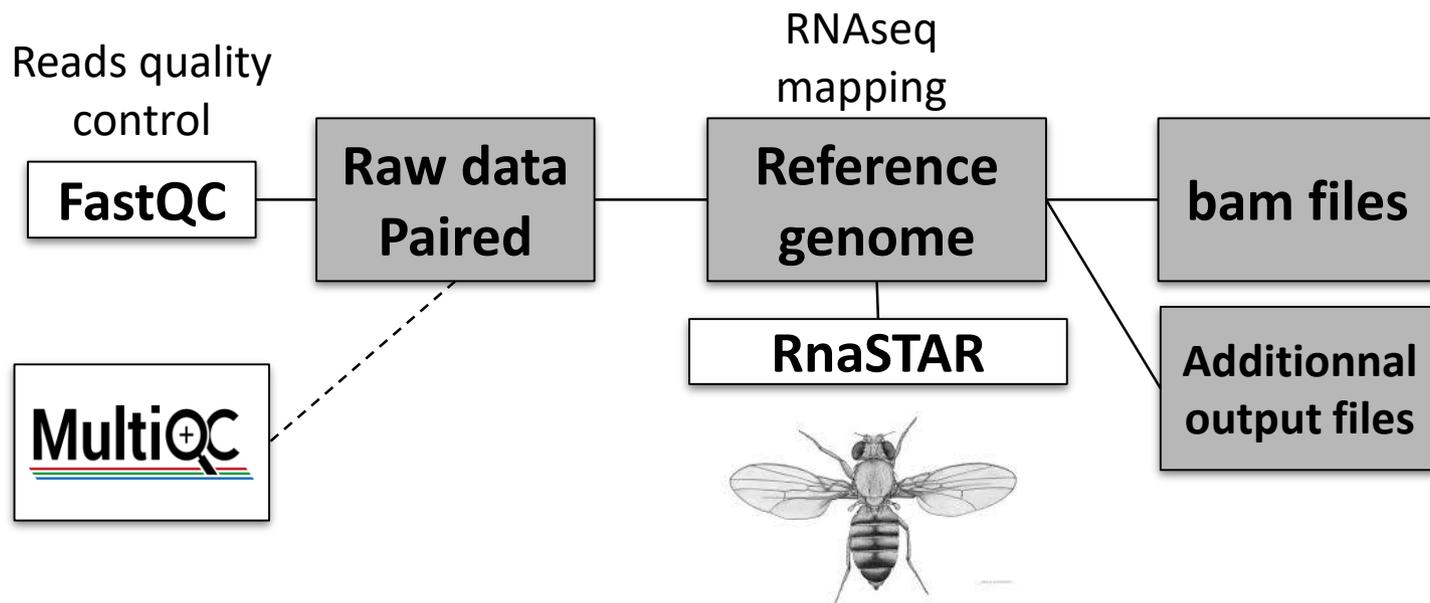
Case of *Drosophila melanogaster* (dm)



Part 1:

You have sequencing data from your favorite species. You want to check the quality of your sequences and to map on the reference genome.

Case of *Drosophila melanogaster* (dm)





Part 1:

You have sequencing data from your favorite species. You want to check the quality of your sequences and to map on the reference genome.

Case of *Drosophila melanogaster* (dm)

1. Look into the description of the tool FastQC, what is its purpose ?
2. Does it takes compressed files ? ".gz"
3. Which encoding of the file (in FastQC results, basic statistics) ? => edit if needed the datatype of your reads files
4. Look at mapping tools, map your reads on genome with RNA STAR.
5. What is the default parameter of the option "Maximum ratio of mismatches to mapped length" in RNA STAR ?
6. How to map on another reference genome or assembly release ?



Part 2:

You have new sequences that you want to compare with the gene and protein databases from your favorite species (BLAST).

Case of *Drosophila melanogaster* (dm)

1. Create new history
2. Import CDS and peptide sequences databases
3. Import query sequences
4. Make BLAST databases
5. Run BLAST against the CDS database
6. Run BLAST against the protein database



Part 2:

You have new sequences that you want to compare with the gene and protein databases from your favorite species (BLAST).

Case of *Drosophila melanogaster* (dm)

1. Create new history
2. Import CDS and peptide sequences databases
3. Import query sequences
4. Make BLAST databases *Blast --> NCBI BLAST+ makeblastdb*
5. Run BLAST against the CDS database
6. Run BLAST against the protein database



Part 2:

You have new sequences that you want to compare with the gene and protein databases from your favorite species (BLAST).

Case of *Drosophila melanogaster* (dm)

1. Create new history
2. Import CDS and peptide sequences databases
3. Import query sequences
4. Make BLAST databases *Blast --> NCBI BLAST+ makeblastdb*
5. Run BLAST against the CDS database *Blast --> NCBI BLAST+ blastn*
6. Run BLAST against the protein database



Part 2:

You have new sequences that you want to compare with the gene and protein databases from your favorite species (BLAST).

Case of *Drosophila melanogaster* (dm)

1. Create new history
2. Import CDS and peptide sequences databases
3. Import query sequences
4. Make BLAST databases *Blast --> NCBI BLAST+ makeblastdb*
5. Run BLAST against the CDS database *Blast --> NCBI BLAST+ blastn*
6. Run BLAST against the protein database *Blast --> NCBI BLAST+ blastx*

DATASET COLLECTION

- Problematic: you have a large numbers of datasets to send through the same analysis

- Problematic: you have a large numbers of datasets to send through the same analysis
- Solution 1: select multiple datasets as input

Dataset collection

Select multiple datasets as input

Galaxy Analyze Data Workflow Shared Data Visualization Help User Using 0%

Workflow: Find exons with highest number of SNPs

Run workflow

History Options
Send results to a new history
Yes No

1: Input dataset
2: Exons

2: Input dataset
4: Repeats
3: SNPs
Multiple datasets Exons

This is a batch mode input field. Separate jobs will be triggered for each dataset selection.
Batch options:

3: Join the intervals of two datasets side-by-side (Galaxy Version 1.0.0)

4: Group data by a column and perform aggregate operation on other columns. (Galaxy Version 2.1.1)

History
search datasets
Galaxy initiation - multiple datasets
3 shown, 1 deleted
10.41 MB
4: Repeats
3: SNPs
2: Exons

Dataset collection

Select multiple datasets as input

The screenshot displays the Galaxy web interface. At the top, the navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. A 'Using 0%' indicator is visible in the top right corner. On the left side, there is a 'Tools' panel with a search bar and a list of tool categories such as 'Get Data', 'Send Data', 'Text Manipulation', and 'NGS: DeepTools'. The main central area features a green notification box with a checkmark icon, stating: 'Successfully invoked workflow **Find exons with highest number of SNPs 2 times.** You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.' On the right side, the 'History' panel shows a search bar and a list of datasets. The top entry is 'Galaxy initiation - multiple datasets' with a size of 11.68 MB. Below it, several datasets are listed, each with a green background and a red border: '14: Top exon genetic location', '13: Top exons', '12: Top exon genetic location', and '10: Top exons'. Other datasets visible include '4: Repeats', '3: SNPs', and '2: Exons'. Each dataset entry includes an eye icon, a pencil icon, and a close icon.

- Problematic: you have a large numbers of datasets to send through the same analysis
- Solution 1: select multiple datasets as input
- Solution 2: create a dataset collection

- Problematic: you have a large numbers of datasets to send through the same analysis
- Solution 1: select multiple datasets as input
- Solution 2: create a dataset collection
 - Dataset list: set of files of the same type
 - Dataset pairs: pairs of read files (forward, reverse)
 - List of dataset pairs

- Problematic: you have a large numbers of datasets to send through the same analysis
- Solution 1: select multiple datasets as input
- Solution 2: create a dataset collection
 - Dataset list: set of files of the same type
 - Dataset pairs: pairs of read files (forward, reverse)
 - List of dataset pairs
- Galaxy runs the tool automatically on each dataset in the collection using the same settings

Dataset collection

Create a dataset collection

The screenshot displays the Galaxy/ABiMS web interface. The top navigation bar includes 'Galaxy / ABiMS', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. A 'Using 0%' indicator is visible in the top right. The left sidebar contains a 'tools' section with a search bar and a list of tool categories: 'Get Data', 'Send Data', 'Collection Operations', 'COMMON TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Convert Formats', 'Extract Features', 'Fetch Sequences', 'Statistics', 'Graph/Display Data', 'Fasta Fastq Manipulation', 'COMMON NGS TOOLS', 'NGS:Samtools', 'NGS:Mapping', 'NGS:Bedtools', 'NGS:Picard Tools', 'SEARCHING TOOLS', and 'Diamond'. The main content area features a green 'Welcome to galaxy3.sb-roscoff.fr' message, an information box with contact details for support.abims@sb-roscoff.fr, the ABiMS logo (Analyses and Bioinformatics for Marine Science), and links to 'Changelog' and 'Tutorials'. The right sidebar shows a 'History' section with a search bar and a collection named 'Galaxy initiation - collection' (3 shown, 11.32 MB). The collection items are: '3: Repeats' (checked), '2: SNPs' (checked), and '1: Exons' (unchecked). A tooltip 'Operations on multiple datasets' is visible over the checked items. The bottom status bar shows 'javascriptvoid(0);'.

Dataset collection

Create a dataset collection

The screenshot displays the Galaxy / ABiMS web interface. The top navigation bar includes 'Galaxy / ABiMS', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', 'User', and a 'Using 0%' indicator. The left sidebar contains a 'Tools' section with a search bar and various tool categories like 'Get Data', 'Send Data', 'Collection Operations', 'COMMON TOOLS', 'COMMON NGS TOOLS', and 'SEARCHING TOOLS'. The main content area features a green welcome banner for 'galaxy3.sb-roscoff.fr', an information box with contact details, and the ABiMS logo with the tagline 'Analyses and Bioinformatics for Marine Science'. Below the logo are links for 'Changelog' and 'Tutorials', followed by a paragraph of text describing the Galaxy platform. The right sidebar shows a 'History' section with a search bar, a 'Galaxy initiation - collection' section showing '3 shown' and '11.32 MB', and a context menu with options: 'Hide datasets', 'Unhide datasets', 'Delete datasets', 'Undelete datasets', 'Permanently delete datasets', 'Build Dataset List', 'Build Dataset Pair', and 'Build List of Dataset Pairs'. The 'Build Dataset List' option is highlighted with a red border.

Dataset collection

Create a dataset collection

The screenshot shows the Galaxy ABiMS interface with a dialog box titled "Create a collection from a list of datasets". The dialog box contains the following elements:

- Header:** "Create a collection from a list of datasets"
- Description:** "Collections of datasets are permanent, ordered lists of datasets that can be passed to tools and workfl... [More help](#)"
- Start over:** A link to "Start over"
- Dataset Selection:** Two input fields with "Repeats" and "SNPs" selected. Each field has a "Discard" button to its right.
- Name:** A text input field containing "Collection of different features", which is highlighted with a red border.
- Buttons:** A "Cancel" button on the bottom left and a "Create list" button on the bottom right, also highlighted with a red border.

The background interface shows the Galaxy navigation menu on the left, including sections like "Tools", "COMMON TOOL", "COMMON NGS TOOLS", and "SEARCHING TOOLS". The top navigation bar includes "Galaxy / ABiMS", "Analyze Data", "Workflow", "Shared Data", "Visualization", "Admin", "Help", and "User". The top right corner shows "Using 0%".

Dataset collection

Create a dataset collection

The screenshot displays the Galaxy / ABiMS web interface. The top navigation bar includes 'Galaxy / ABiMS', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', 'User', and a 'Using 0%' indicator. The left sidebar contains a 'Tools' section with a search bar and a list of tool categories: 'Get Data', 'Send Data', 'Collection Operations', 'COMMON TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Convert Formats', 'Extract Features', 'Fetch Sequences', 'Statistics', 'Graph/Display Data', 'Fasta Fastq Manipulation', 'COMMON NGS TOOLS', 'NGS:Samtools', 'NGS:Mapping', 'NGS:Bedtools', 'NGS:Picard Tools', 'SEARCHING TOOLS', and 'Diamond'. The main content area features a green welcome banner for 'galaxy3.sb-roscoff.fr', an information box with contact details for support.abims@sb-roscoff.fr, the ABiMS logo (Analyses and Bioinformatics for Marine Science), and links to 'Changelog' and 'Tutorials'. Below these is a paragraph describing Galaxy as an open, web-based platform for data-intensive biomedical research, supported by various institutions. The right sidebar shows a 'History' section with a search bar, 'Galaxy initiation - collection' (4 shown, 11.32 MB), and a list of dataset collections. The collection '4: Collection of different features' is highlighted with a red border and contains 'a list of datasets'. Other collections listed are '3: Repeats', '2: SNPs', and '1: Exons'.

Galaxy / ABiMS Analyze Data Workflow Shared Data Visualization Admin Help User Using 0%

Tools search tools

Get Data Send Data Collection Operations

COMMON TOOLS Text Manipulation Filter and Sort Join, Subtract and Group Convert Formats Extract Features Fetch Sequences Statistics Graph/Display Data Fasta Fastq Manipulation

COMMON NGS TOOLS NGS:Samtools NGS:Mapping NGS:Bedtools NGS:Picard Tools

SEARCHING TOOLS Diamond

✓ Welcome to galaxy3.sb-roscoff.fr

Information For any question or request for tools or account, send an email at support.abims@sb-roscoff.fr

ABiMS Analyses and Bioinformatics for Marine Science

Station Biologique Roscoff

Changelog Tutorials

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 search datasets

Galaxy initiation - collection 4 shown 11.32 MB

All None For all selected...

4: Collection of different features a list of datasets

3: Repeats

2: SNPs

1: Exons

Dataset collection

Create a dataset collection

The screenshot displays the Galaxy / ABiMS web interface. The top navigation bar includes links for 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User', along with a 'Using 0%' indicator. The left sidebar lists various tool categories such as 'Get Data', 'Send Data', 'Collection Operations', 'COMMON TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Convert Formats', 'Extract Features', 'Fetch Sequences', 'Statistics', 'Graph/Display Data', 'Fasta Fastq Manipulation', 'COMMON NGS TOOLS', 'NGS:Samtools', 'NGS:Mapping', 'NGS:Bedtools', 'NGS:Picard Tools', 'SEARCHING TOOLS', and 'Diamond'. The main content area features a green 'Welcome to galaxy3.sb-roscoff.fr' message, an 'Information' box with contact details, the ABiMS logo (Analyses and Bioinformatics for Marine Science), and links to 'Changelog' and 'Tutorials'. A paragraph of text describes Galaxy as an open, web-based platform for data-intensive biomedical research, supported by various institutions. The right-hand panel shows a 'History' section with a red-bordered box highlighting a collection titled 'Collection of different features' (a list of datasets), which includes sub-items 'Repeats' and 'SNPs'.

Dataset collection

Use a collection as input

BED-to-GFF converter (Galaxy Version 2.0.0) Options

Convert this dataset

5: Collection of different features

Dataset collection

Execute

What it does

This tool converts data from BED format to GFF format (scroll down for format description).

Example

The following data in BED format:

```
chr28 346187 388197 BC114771 0 + 346187 388197 0
```

Will be converted to GFF (note that the start coordinate is incremented by 1):

```
chr28 bed2gff mRNA 346188 388197 0 + . mRNA BC1147
chr28 bed2gff exon 346188 346331 0 + . exon BC1147
chr28 bed2gff exon 370283 370363 0 + . exon BC1147
chr28 bed2gff exon 372378 372492 0 + . exon BC1147
chr28 bed2gff exon 377194 377256 0 + . exon BC1147
chr28 bed2gff exon 378319 378473 0 + . exon BC1147
chr28 bed2gff exon 379722 379817 0 + . exon BC1147
chr28 bed2gff exon 383182 383315 0 + . exon BC1147
chr28 bed2gff exon 387981 388085 0 + . exon BC1147
```

History

search datasets

Galaxy initiation - collection
4 shown, 5 deleted, 6 hidden

11.68 MB

5: Collection of different features
a list of 2 datasets

4: Repeats

2: SNPs

1: Exons

Dataset collection

Use a collection as input

The screenshot shows the Galaxy web interface. At the top, there is a navigation bar with 'Galaxy' logo and menu items: 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. A 'Using 0%' indicator is in the top right. On the left, a 'tools' sidebar contains a search bar and various tool categories like 'Get Data', 'Send Data', 'Lift-Over', 'Text Manipulation', 'Datamash', and 'Convert Formats'. The main content area features a green notification box with a checkmark icon, stating: '2 jobs have been successfully added to the queue - resulting in the following datasets: 16: BED-to-GFF on data 4, 17: BED-to-GFF on data 2. You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.' On the right, the 'History' panel shows a search bar and a list of jobs. A red box highlights the job '18: BED-to-GFF on collection 5', which is described as 'a list of datasets'. Below it are jobs '17: BED-to-GFF on data 2' and '16: BED-to-GFF on data 4'. Further down are '5: Collection of different features', '4: Repeats', '2: SNPs', and '1: Exons'. Each job entry includes an eye icon, a pencil icon, and a close 'x' icon.

Hands-on
COLLECTION





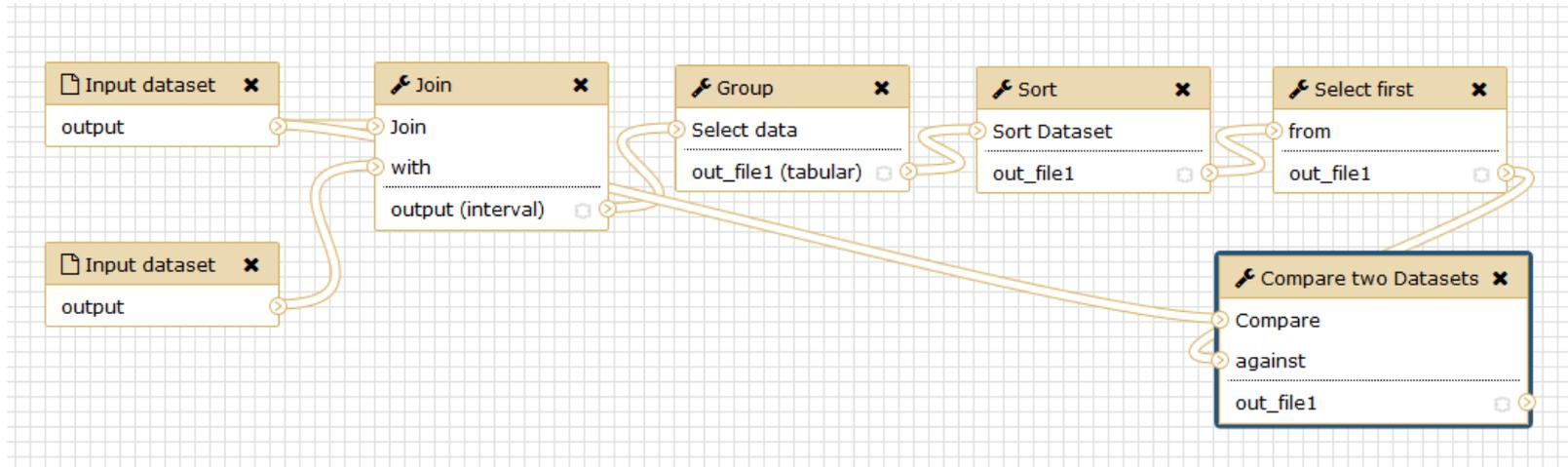
Part 2:

You have new sequences that you want to compare with the gene and protein databases from your favorite species (BLAST).

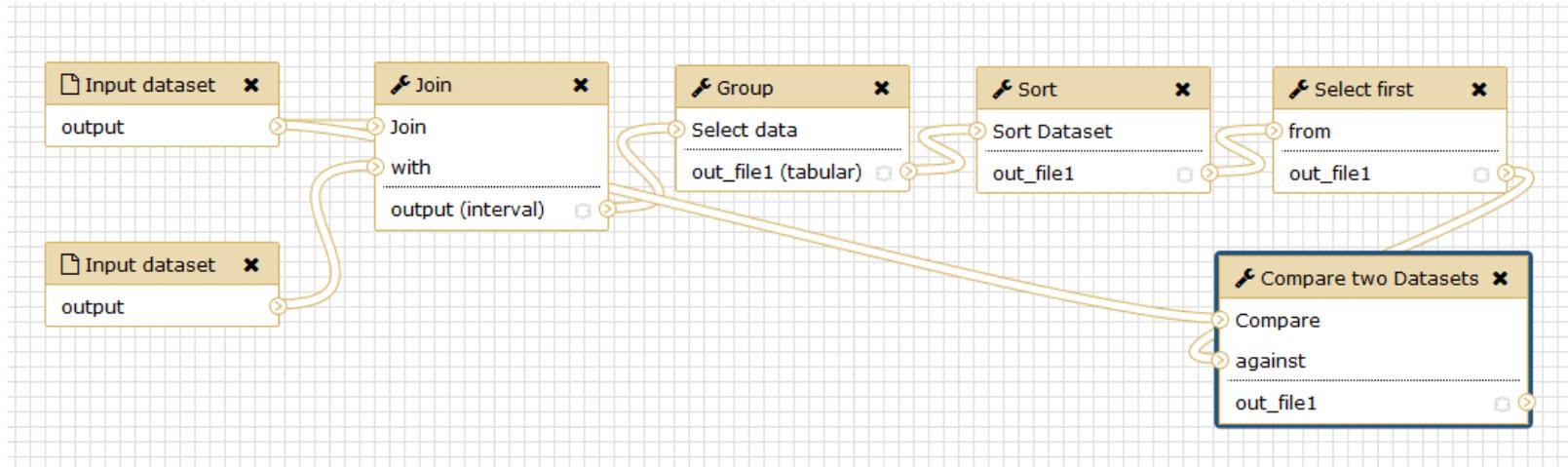
Case of *Drosophila melanogaster* (dm)

1. Create new history
2. Import CDS and peptide sequences databases
3. Import query sequences
4. Make BLAST databases
5. Run BLAST against the CDS database
6. Run BLAST against the protein database
7. Create a dataset list with all the query sequences
8. Run BLAST against the CDS database on the dataset list

WORKFLOW



- A workflow is a sequence of tool operations and parameters
- A workflow is built to be replayed (more or less strict)

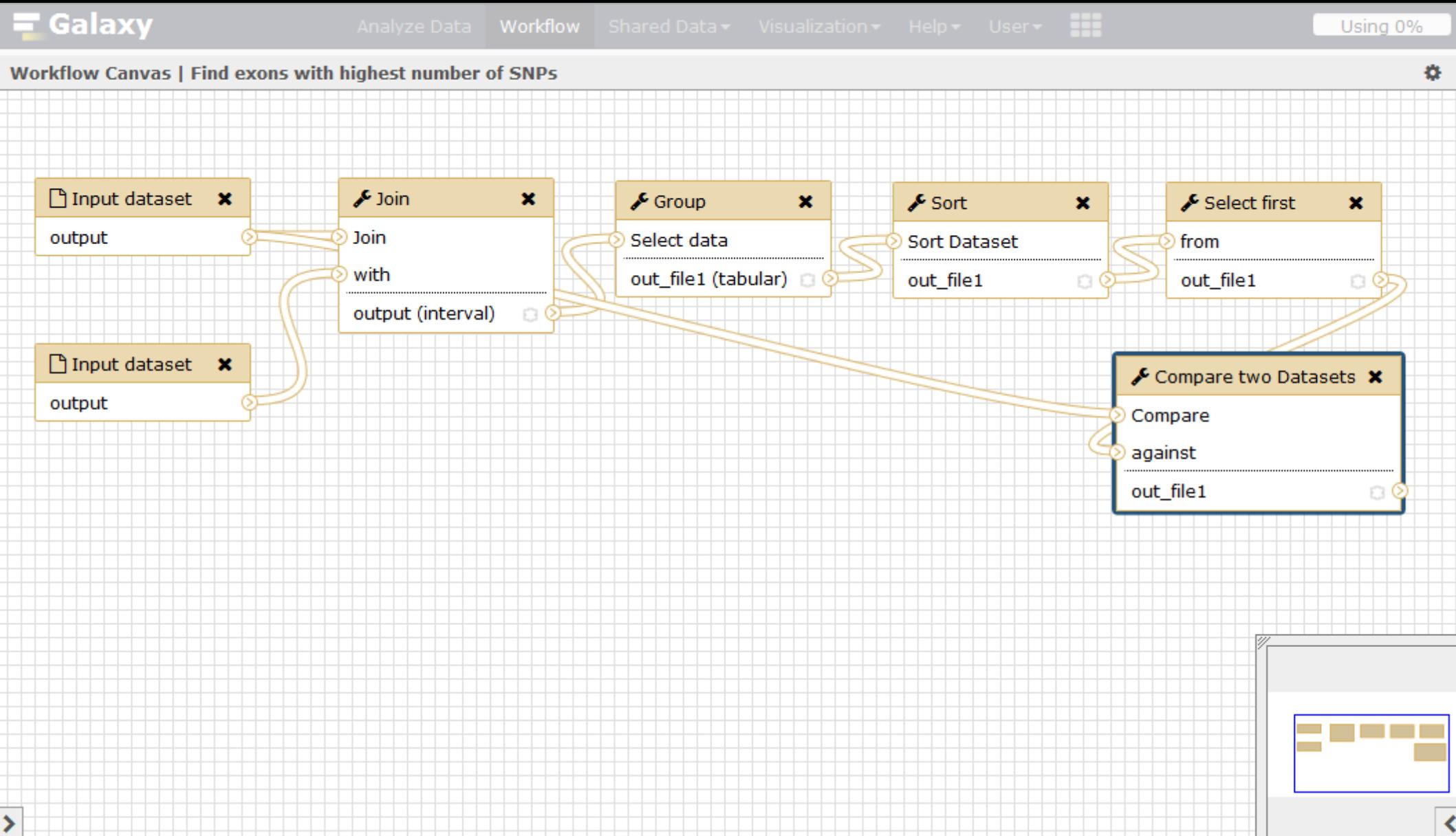


Why would you want to create workflows?

- Re-run the same analysis on different input data sets
- Change parameters before re-running a similar analysis
- Make use of the workflow job scheduling (jobs are submitted as soon as their inputs are ready)
- Share workflows for publication and with the community

Workflow

Our workflow with Galaxy



Workflow

From history

Galaxy Analyze Data Workflow Shared Data Visualization Help User Using 0%

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 'Galaxy initiation'
Create Workflow Check all Uncheck all

Tool	History items created
UCSC Main <i>This tool cannot be used in workflows</i>	1 Exons <input checked="" type="checkbox"/> Treat as input dataset Exons
UCSC Main <i>This tool cannot be used in workflows</i>	2 SNPs <input checked="" type="checkbox"/> Treat as input dataset SNPs
Join <input checked="" type="checkbox"/> Include "Join" in workflow	3 Join on data 2 and data 1
Group <input checked="" type="checkbox"/> Include "Group" in workflow	5 Group on data 3
Sort <input checked="" type="checkbox"/> Include "Sort" in workflow	6 Sort on data 5

https://usegalaxy.org/workflow/build_from_current_history

History History Lists Saved Histories Histories Shared with Me CURRENT HISTORY Create New Copy History Share or Publish Show Structure **7: Extract Workflow** 6: Delete 6: Delete Permanently DATASET ACTIONS Copy Datasets Dataset Security Resume Paused Jobs Collapse Expanded Datasets Unhide Hidden Datasets Delete Hidden Datasets Purge Deleted Datasets DOWNLOADS Export Tool Citations Export History to File OTHER ACTIONS Import from File

Workflow

From history

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
Find exons with the highest SNPs

Create Workflow Check all Uncheck all

Tool	History items created
UCSC Main <i>This tool cannot be used in workflows</i>	1 Exons <input checked="" type="checkbox"/> Treat as input dataset Exons
UCSC Main <i>This tool cannot be used in workflows</i>	2 SNPs <input checked="" type="checkbox"/> Treat as input dataset SNPs
Join <input checked="" type="checkbox"/> Include "Join" in workflow	3 Join on data 2 and data 1
Group <input checked="" type="checkbox"/> Include "Group" in workflow	5 Group on data 3
Sort <input checked="" type="checkbox"/> Include "Sort" in workflow	6 Sort on data 5

History
search datasets

Galaxy initiation
7 shown, 1 deleted, 1 hidden
8.77 MB

8: Compare two Datasets on data 7 and data 1

7: Select first on data 6

6: Sort on data 5

5: Group on data 3

3: Join on data 2 and data 1

2: SNPs

1: Exons

Workflow

From history

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
Find exons with the highest SNPs

Create Workflow Check all Uncheck all

Tool	History items created
UCSC Main <i>This tool cannot be used in workflows</i>	1 Exons <input checked="" type="checkbox"/> Treat as input dataset Exons
UCSC Main <i>This tool cannot be used in workflows</i>	2 SNPs <input checked="" type="checkbox"/> Treat as input dataset SNPs
Join <input checked="" type="checkbox"/> Include "Join" in workflow	3 Join on data 2 and data 1
Group <input checked="" type="checkbox"/> Include "Group" in workflow	5 Group on data 3
Sort <input checked="" type="checkbox"/> Include "Sort" in workflow	6 Sort on data 5

History
search datasets

Galaxy initiation
7 shown, 1 deleted, 1 hidden
8.77 MB

8: Compare two Datasets on data 7 and data 1

7: Select first on data 6

6: Sort on data 5

5: Group on data 3

3: Join on data 2 and data 1

2: SNPs

1: Exons

Workflow

Workflow manager

Galaxy Analyze Data **Workflow** Shared Data Visualization Help User Using 0%

Your workflows

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

Name	# of Steps
Find exons with highest number of SNPs	7
Convert to tab (imported from API)	2
imported: ChIP-seq workflow	3

Workflows shared with you by others

No workflows have been shared with you.

Other options

[Configure your workflow menu](#)

Workflow

Workflow manager

Galaxy Analyze Data Workflow Shared Data Visualization Help User Using 0%

Your workflows

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

Name	# of Steps
Find exons with highest number of SNPs 	7
Convert to tab (imported from ...)	2
imported: ChIP-seq workflow	3

Workflows shared with others
No workflows have been shared with others.

Other options
[Configure your workflow menu](#)

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

<https://usegalaxy.org/workflow/editor?id=17b7895387cc2214>

Workflow

Edit a workflow: add tags and annotation

The screenshot displays the Galaxy workflow editor interface. The top navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. A 'Using 0%' indicator is visible in the top right. The main area is divided into three sections: 'Tools', 'Workflow Canvas', and 'Details'.

Tools: A search bar labeled 'search tools' is at the top. Below it, a list of tool categories is provided, including 'Inputs', 'Get Data', 'Send Data', 'Lift-Over', 'Text Manipulation', 'Datamash', 'Convert Formats', 'Filter and Sort', 'Join, Subtract and Group', 'Fetch Alignments/Sequences', 'NGS: QC and manipulation', 'NGS: DeepTools', 'NGS: Mapping', 'NGS: RNA Analysis', 'NGS: SAMtools', 'NGS: BamTools', 'NGS: Picard', 'NGS: VCF Manipulation', 'NGS: Peak Calling', 'NGS: Variant Analysis', 'NGS: RNA Structure', and 'NGS: Du Novo'.

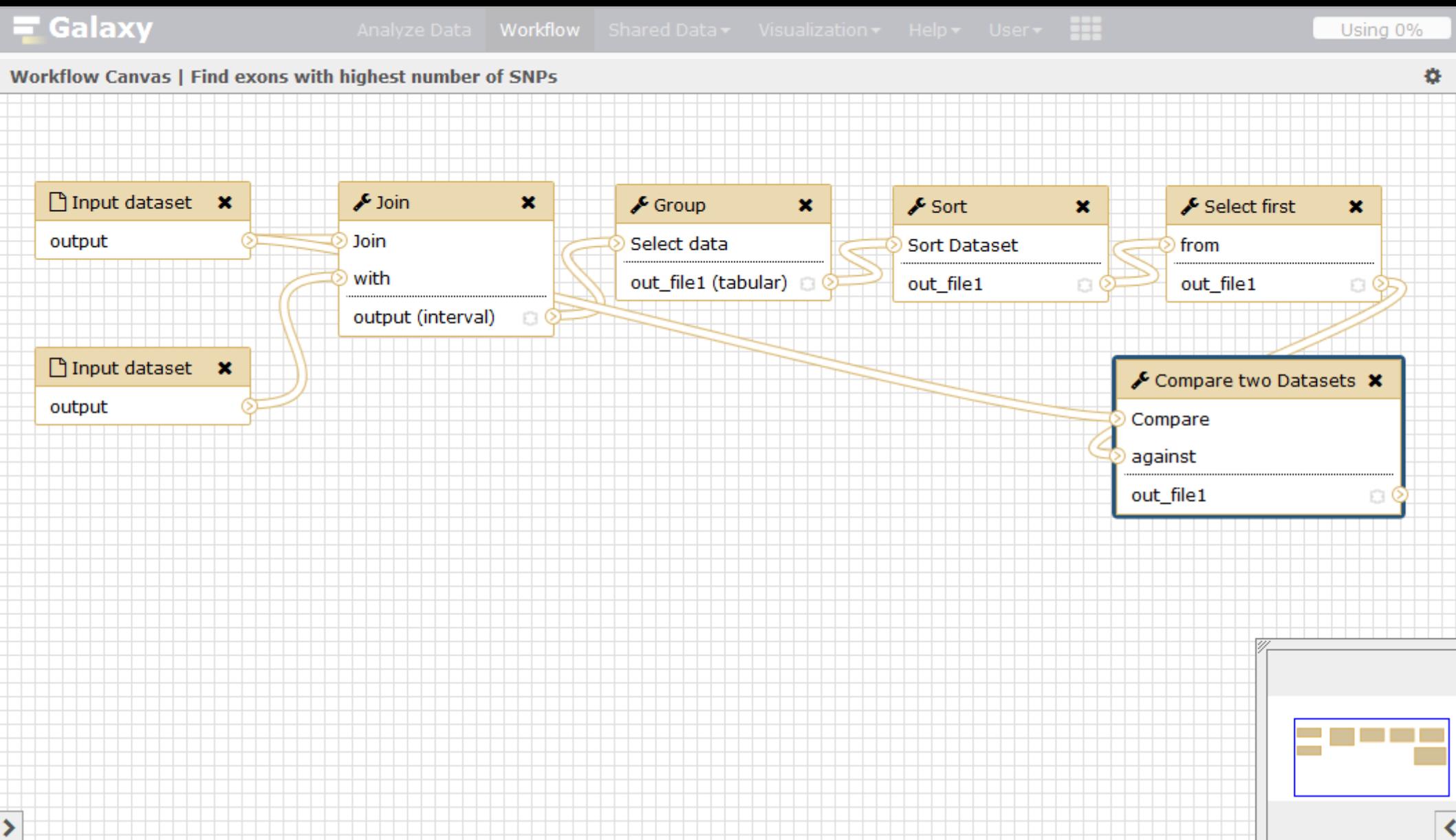
Workflow Canvas: The canvas is titled 'Workflow Canvas | Find exons with highest number of SNPs'. It features a grid background with three main workflow steps: two 'Input dataset' blocks (each with an 'output' field) and a 'Join' block. The 'Join' block has a 'with' section containing 'output (interval)'. The output of the 'Join' block is connected to a 'Group' block, which has a 'Select data' section containing 'out_file1 (tab)'. A small preview window at the bottom right shows a grid of colored squares.

Details: The 'Details' panel is titled 'Edit Workflow Attributes' and is highlighted with a red border. It contains the following information:

- Name:** Find exons with highest number of SNPs
- Tags:** A section with a plus icon and a text box containing 'out_file1 (tab)'. Below it, a description reads: 'Apply tags to make it easy to search for and find items with the same tag.'
- Annotation / Notes:** A section with a text box containing the text: 'Describe or add notes to workflow. Add an annotation or notes to a workflow; annotations are available when a workflow is viewed.'

Workflow

Edit a workflow



Workflow

Edit a workflow: drag and drop

The screenshot displays the Galaxy Workflow Canvas interface. The title bar reads "Workflow Canvas | Find exons with highest number of SNPs". The workflow consists of the following steps:

- Input dataset**: Two input datasets, each with an "output" port.
- Join**: A "Join" tool that takes two inputs and produces an "output (interval)".
- Group**: A "Group" tool that takes the output from the "Join" step and produces "out_file1 (tabular)".
- Sort**: A "Sort" tool that takes the output from the "Group" step and produces "out_file1".
- Select first**: A "Select first" tool that takes the output from the "Sort" step and produces "out_file1".
- Compare two Datasets**: A "Compare two Datasets" tool that takes the output from the "Select first" step and produces "out_file1".

The workflow is visualized as a sequence of connected boxes on a grid background. The "Compare two Datasets" tool is highlighted with a blue border. A small preview window in the bottom right corner shows a simplified view of the workflow.

Workflow

Edit a workflow: delete a noodle

The screenshot shows the Galaxy Workflow Canvas interface. The title bar reads "Workflow Canvas | Find exons with highest number of SNPs". The workflow consists of the following tools and connections:

- Two "Input dataset" tools (output) connected to the "Join" tool.
- The "Join" tool (with "output (interval)" selected) connects to the "Group" tool.
- The "Group" tool (with "out_file1 (tabular)" selected) connects to the "Sort" tool.
- The "Sort" tool (with "out_file1" selected) connects to the "Select first" tool.
- The "Select first" tool (with "out_file1" selected) connects to the "Compare two Datasets" tool.
- The "Compare two Datasets" tool has two "Compare" options. The top "Compare" option is highlighted with a red box, indicating it is the current selection for editing.

The "Compare two Datasets" tool configuration is as follows:

- Compare
- against
- out_file1

Workflow

Edit a workflow: delete a noodle

The screenshot shows the Galaxy Workflow Canvas interface. The title bar reads "Workflow Canvas | Find exons with highest number of SNPs". The main workspace contains a workflow with the following steps:

- Input dataset** (output)
- Input dataset** (output)
- Join** (Join with output (interval))
- Group** (Select data out_file1 (tabular))
- Sort** (Sort Dataset out_file1)
- Select first** (from out_file1)
- Compare two Datasets** (Compare against out_file1)

The workflow is connected as follows: the two input datasets feed into the Join step. The output of the Join step feeds into the Group step. The output of the Group step feeds into the Sort step. The output of the Sort step feeds into the Select first step. The Compare two Datasets step is currently disconnected from the main workflow.

Workflow

Edit a workflow: add a tool

The screenshot displays the Galaxy workflow editor interface. The top navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. A 'Using 0%' indicator is visible in the top right corner. The left sidebar contains a 'Tools' panel with a search bar and a list of tool categories: 'Get Data', 'Send Data', 'Lift-Over', and 'Text Manipulation'. The 'Text Manipulation' section is expanded, showing a list of tools, with 'Merge Columns together' highlighted by a red box. The main 'Workflow Canvas' is titled 'Find exons with highest number of SNPs' and features a grid background. The workflow consists of several tools connected by arrows: two 'Input dataset' tools feeding into a 'Join' tool; the 'Join' tool feeding into a 'Group' tool; the 'Group' tool feeding into a 'Sort' tool; and a 'Merge Columns' tool (highlighted with a red box) positioned below the 'Join' tool. The 'Merge Columns' tool is currently selected, showing its configuration options: 'Select data' and 'out_file1 (tabular)'. Below it, a 'Compare two Datasets' tool is also visible, with configuration options 'Compare', 'against', and 'out_file1'. A small preview window in the bottom right corner shows a grid of tool icons.

Workflow

Edit a workflow: add a noodle

The screenshot displays the Galaxy workflow editor interface. On the left, there is a 'Tools' sidebar with a search bar and a list of tool categories: 'Get Data', 'Send Data', 'Lift-Over', and 'Text Manipulation'. The 'Text Manipulation' section is expanded, showing a list of tools such as 'UniProt ID mapping and retrieval', 'Compute an expression on every row', 'Concatenate datasets tail-to-head', 'Add column to an existing dataset', 'Concatenate datasets tail-to-head (cat)', 'tac reverse a file (reverse cat)', 'Condense consecutive characters', 'Cut columns from a table', 'Convert delimiters to TAB', 'Merge Columns together', and 'Remove beginning of a file'. The main area is the 'Workflow Canvas' titled 'Find exons with highest number of SNPs'. It contains several tools connected by lines representing data flow. The tools include: 'Input dataset' (two instances), 'Join', 'Merge Columns', 'Group', 'Sort', and 'Compare two Datasets'. A red box highlights the output of the top 'Input dataset' tool, with a red arrow pointing to the 'Join' tool. A green line connects the output of the 'Join' tool to the 'Merge Columns' tool. The 'Merge Columns' tool has a 'Select data' section with 'out_file1 (tabular)'. The 'Group' tool has a 'Select data' section with 'out_file1 (tabular)'. The 'Sort' tool has a 'Sort Dataset' section with 'out_file1'. The 'Compare two Datasets' tool has a 'Compare' section with 'against' and 'out_file1'. A small preview window in the bottom right corner shows a grid of data points.

Workflow

Edit a workflow: hide intermediate steps

The screenshot shows the Galaxy Workflow Canvas interface. The title bar reads "Workflow Canvas | Find exons with highest number of SNPs". The workflow consists of the following steps:

- Join**: Takes two "Input dataset" steps as input. The output is labeled "output (interval)".
- Group**: Takes the output from the "Join" step as input. The output is labeled "out_file1 (tabular)".
- Sort**: Takes the output from the "Group" step as input. The output is labeled "out_file1".
- Select first**: Takes the output from the "Sort" step as input. The output is labeled "out_file1".
- Compare two Datasets**: Takes the output from the "Select first" step as input. The output is labeled "out_file1".

A tooltip is displayed over the "Compare two Datasets" step, containing the following text:

Mark dataset as a workflow output. All unmarked datasets will be hidden.

Workflow

Edit a workflow: set or release a parameter

The screenshot displays the Galaxy Workflow Canvas interface. The workflow is titled "Find exons with highest number of SNPs" and consists of the following steps:

- Group**: Select data (out_file1 (tabular))
- Sort**: Sort Dataset (out_file1)
- Select first**: from (out_file1) - This step is highlighted with a red box.
- Compare two Datasets**: Compare against (out_file1)

The configuration panel for the "Select first" tool is open on the right side of the interface. It includes the following fields:

- Select first lines from a dataset (Galaxy Version 1.0.0)**: A dropdown menu.
- Select first**: A checked checkbox.
- 5**: A text input field for the number of lines.
- lines**: A label for the input field.
- from**: A label for the data input.
- Data input 'input' (txt)**: A text input field.
- Annotation / Notes**: A large text area for adding notes.
- Add an annotation or note for this step. It will be shown with the workflow.**: A descriptive text for the annotation field.
- Email notification**: A section with "Yes" and "No" radio buttons.
- An email notification will be sent when the job has completed.**: A descriptive text for the email notification field.
- Output cleanup**: A section with "Yes" and "No" radio buttons.

Workflow

Edit a workflow: set or release a parameter

The screenshot displays the Galaxy Workflow Canvas for a workflow titled "Find exons with highest number of SNPs". The workflow is composed of the following steps:

- Group**: Select data (out_file1 (tabular))
- Sort**: Sort Dataset (out_file1)
- Select first**: from (out_file1) - This step is highlighted with a blue border.
- Compare two Datasets**: Compare against (out_file1)
- Select first**: (out_file1)

The configuration panel for the selected "Select first" step is visible on the right, showing the following options:

- Select first** (highlighted with a red box)
- Set at Runtime** (selected)
- lines**
- from**: Data input 'input' (txt)
- Annotation / Notes**: (empty text area)
- Email notification**: Yes No
- Output cleanup**: Yes No

Workflow

Edit a workflow: rename the outputs

The screenshot shows the Galaxy workflow editor interface. The top navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The workflow canvas is titled 'Workflow Canvas | Find exons with highest number of SNPs'. The workflow consists of four steps:

- Group**: Select data, out_file1 (tabular)
- Sort**: Sort Dataset, out_file1
- Select first**: from, out_file1 (highlighted with a blue border)
- Compare two Datasets**: Compare, against, out_file1 (highlighted with a blue border)

The 'Details' panel on the right shows the configuration for the selected step. The 'Rename dataset' option is highlighted with a red border, with the text 'Top exons' entered in the input field.

Details

step. It will be shown with the workflow.

Email notification

An email notification will be sent when the job has completed.

Output cleanup

Upon completion of this step, delete non-starred outputs from completed workflow steps if they are no longer required as inputs.

Configure Output: 'out_file1'

Label

This will provide a short name to describe the output - this must be unique across workflows.

Rename dataset

This action will rename the output dataset. Click [here](#) for more information. Valid inputs are: **input**.

Workflow

Save

Galaxy Analyze Data Workflow Shared Data Visualization Help User Using 0%

Workflow Canvas | Find exons with highest number of SNPs

Group Select data out_file1 (tabular)

Sort Sort Dataset out_file1

Select first from out_file1

Compare two Datasets Compare against out_file1

Save
Save As
Run
Edit Attributes
Auto Re-layout
Close

Details

from
Data input 'input' (txt)

Annotation / Notes

Add an annotation or note for this step. It will be shown with the workflow.

Email notification

Yes No
An email notification will be sent when the job has completed.

Output cleanup

Yes No

Workflow

Run a workflow

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow' (highlighted with a red box), 'Shared Data', 'Visualization', 'Help', 'User', and a 'Using 0%' indicator. Below the navigation bar, the 'Your workflows' section features two buttons: 'Create new workflow' and 'Upload or import workflow'. A table lists workflows with columns for 'Name' and '# of Steps'. The first workflow, 'Find exons with highest number of SNPs', is highlighted with a red box. A context menu is open over this workflow, with the 'Run' option highlighted in blue and also boxed in red. Other menu options include 'Edit', 'Share or Download', 'Copy', 'Rename', 'View', and 'Delete'. Below the table, there are sections for 'Workflows shared by others' (with a message 'No workflows have been shared by others') and 'Other options' (with a button 'Configure your workflow menu'). At the bottom left, a URL is visible: https://usegalaxy.org/root?workflow_id=17b7895387cc2214

Name	# of Steps
Find exons with highest number of SNPs	7
Convert to tab (imported from ...)	2
imported: CHIP-seq workflow	3

Workflow

Run a workflow

The screenshot displays the Galaxy workflow editor interface. The main workspace shows a workflow titled "Workflow: Find exons with highest number of SNPs" with a "Run workflow" button. The workflow steps are:

- History Options:** "Send results to a new history" with "Yes" selected.
- 1: Input dataset:** A dropdown menu showing "2: Exons" (highlighted with a red box).
- 2: Input dataset:** A dropdown menu showing "1: Repeats" (highlighted with a red box).
- 3: Join the intervals of two datasets side-by-side (Galaxy Version 1.0.0)**
- 4: Group data by a column and perform aggregate operation on other columns. (Galaxy Version 2.1.1)**
- 5: Sort data in ascending or descending order (Galaxy Version 1.0.3)**
- 6: Select first lines from a dataset (Galaxy Version 1.0.0)** (highlighted with a red box). Below this step, a text input field contains "20" and the label "lines" is visible.

The left sidebar contains a "Tools" section with a search bar and a list of tool categories including "Get Data", "Send Data", "Text Manipulation", "Datamash", "Convert Formats", "Filter and Sort", "Join, Subtract and Group", "Fetch Alignments/Sequences", "NGS: QC and manipulation", "NGS: DeepTools", "NGS: Mapping", "NGS: RNA Analysis", "NGS: SAMtools", "NGS: BamTools", "NGS: Picard", "NGS: VCF Manipulation", "NGS: Peak Calling", "NGS: Variant Analysis", "NGS: RNA Structure", "NGS: Du Novo", and "NGS: Gemini".

The right sidebar shows the "History" section with a search bar and a list of datasets. The current workflow history includes "Galaxy initiation - workflow" (2 shown, 2.77 MB) and two datasets: "2: Exons" and "1: Repeats", both highlighted in green.

Workflow

Run a workflow

Galaxy Analyze Data Workflow Shared Data Visualization Help User Using 0%

Tools search tools

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Datamash
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- Fetch Alignments/Sequences
- NGS: QC and manipulation
- NGS: DeepTools
- NGS: Mapping
- NGS: RNA Analysis
- NGS: SAMtools
- NGS: BamTools
- NGS: Picard
- NGS: VCF Manipulation
- NGS: Peak Calling
- NGS: Variant Analysis
- NGS: RNA Structure
- NGS: Du Novo
- NGS: Gemini

Workflow: Find exons with highest number of SNPs

Run workflow

History Options

Send results to a new history
Yes No

1: Input dataset
2: Exons

2: Input dataset
1: Repeats

3: Join the intervals of two datasets side-by-side (Galaxy Version 1.0.0)

Join
Output dataset 'output' from step 1
with
Output dataset 'output' from step 2
 with min overlap
Edit

Return
Only records that are joined (INNER JOIN)

History search datasets

Galaxy initiation - workflow
2 shown, 2 deleted, 3 hidden
2.92 MB

2: Exons

1: Repeats

Workflow

Run a workflow

Galaxy Analyze Data Workflow Shared Data Visualization Help User Using 0%

Workflow: Find exons with highest number of SNPs Run workflow

History Options

Send results to a new history
Yes No

1: Input dataset
2: Exons

2: Input dataset
1: Repeats

3: Join the intervals of two datasets side-by-side (Galaxy Version 1.0.0)

4: Group data by a column and perform aggregate operation on other columns. (Galaxy Version 2.1.1)

5: Sort data in ascending or descending order (Galaxy Version 1.0.3)

6: Select first lines from a dataset (Galaxy Version 1.0.0)

Select first
20
lines

History search datasets 2 shown
Galaxy initiation - workflow 2.77 MB
2: Exons
1: Repeats

Workflow

Run a workflow

The screenshot displays the Galaxy web interface. At the top, the navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. A 'Using 0%' indicator is visible in the top right corner. On the left, a 'Tools' sidebar lists various categories such as 'Get Data', 'Send Data', 'Text Manipulation', and 'NGS: QC and manipulation'. The main workspace contains a green notification box with a checkmark icon, stating: 'Successfully invoked workflow **Find exons with highest number of SNPs.** You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.' On the right, the 'History' pane shows a list of workflow steps. Steps 1 through 7 are listed, with steps 3 through 7 highlighted in red. Step 1 is '1: Repeats', step 2 is '2: Exons', step 3 is '3: Join on data 1 and data 2', step 4 is '4: Group on data 3', step 5 is '5: Sort on data 4', step 6 is '6: Top exons', and step 7 is '7: Top exon genetic location'. Each step includes icons for viewing, editing, and deleting.

Galaxy Analyze Data Workflow Shared Data Visualization Help User Using 0%

Tools

search tools

Get Data
Send Data
Lift-Over
Text Manipulation
Datamash
Convert Formats
Filter and Sort
Join, Subtract and Group
Fetch Alignments/Sequences
NGS: QC and manipulation
NGS: DeepTools
NGS: Mapping
NGS: RNA Analysis
NGS: SAMtools
NGS: BamTools
NGS: Picard
NGS: VCF Manipulation
NGS: Peak Calling
NGS: Variant Analysis
NGS: RNA Structure
NGS: Du Novo
NGS: Gemini

Successfully invoked workflow **Find exons with highest number of SNPs.**
You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History

search datasets

Galaxy initiation - workflow
7 shown
2.77 MB

7: Top exon genetic location

6: Top exons

5: Sort on data 4

4: Group on data 3

3: Join on data 1 and data 2

2: Exons

1: Repeats

Workflow

Run a workflow

The screenshot displays the Galaxy web interface. At the top, the navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. A 'Using 0%' indicator is visible in the top right corner.

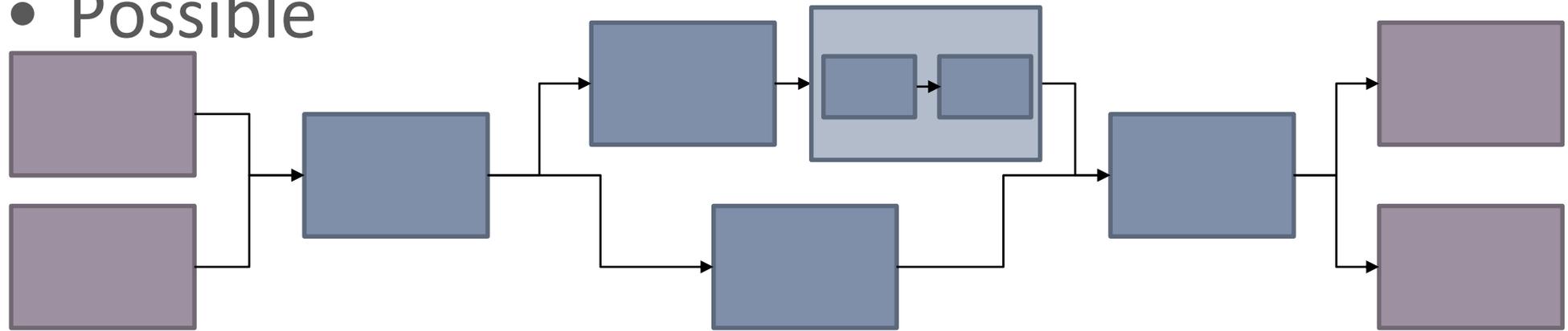
On the left side, there is a 'Tools' panel with a search bar and a list of tool categories such as 'Get Data', 'Send Data', 'Lift-Over', 'Text Manipulation', 'Datamash', 'Convert Formats', 'Filter and Sort', 'Join, Subtract and Group', 'Fetch Alignments/Sequences', and various NGS-related tools.

The main workspace contains a green notification box with a checkmark icon. The text reads: 'Successfully invoked workflow **Find exons with highest number of SNPs.** You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.'

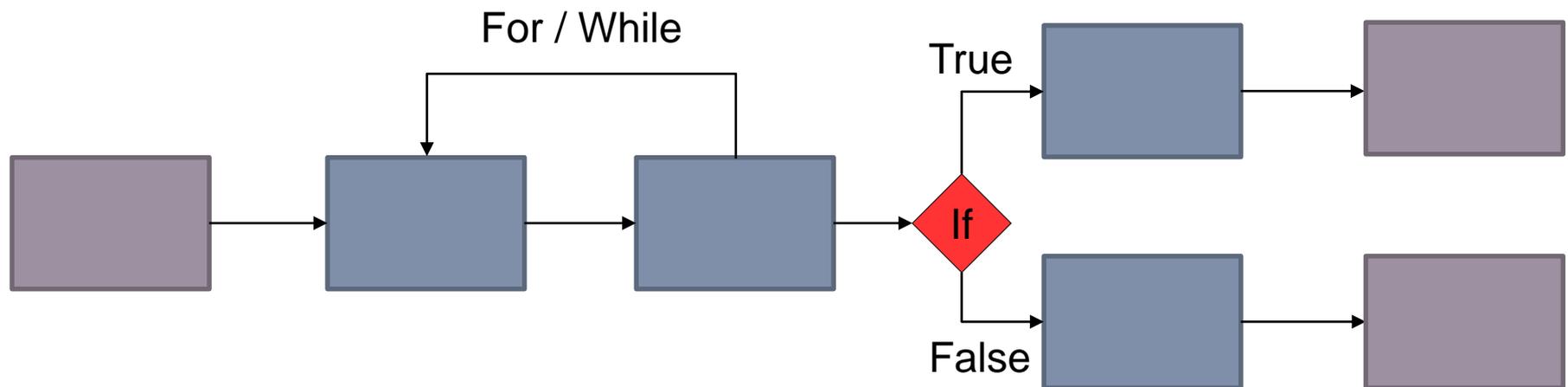
On the right side, the 'History' panel is visible. It features a search bar for datasets and shows a list of workflow jobs. The top job is 'Galaxy initiation - workflow' with 4 shown and 3 hidden. Below it, several jobs are listed with their names and icons for viewing, editing, and deleting:

- 7: Top exon genetic location** (highlighted with a red box)
- 6: Top exons** (highlighted with a red box)
- 2: Exons**
- 1: Repeats**

- Possible



- Impossible (until now)



SHARE

biologist ↔ biologist

- Sharing histories or datasets
 - With or without linked workflow

bioanalyst ↔ biologist

- Sharing workflows
 - Pre-configured parameters
 - With or without release parameters (set at runtime)
 - According to the user-end knowledge

bioinformatician ↔ bioinformatician

- Sharing tools ,scripts and wrappers
 - Toolshed



Share

History

The screenshot displays the Galaxy web interface. The main content area is titled "Saved Histories" and contains a table of saved history items. A context menu is open over the first item, "Galaxy initiation - workflow", with the "Share or Publish" option highlighted. The right sidebar shows a "History" panel with a search bar and a list of datasets.

Name	Datasets	Tags	Sharing	Size on Disk	Created	Last Upd
Galaxy initiation - workflow	4	0 Tags		2.9 MB	~5 hours ago	~4 hours
-		0 Tags		8.8 MB	~10 hours ago	~5 hours
imported: Galaxy 101 (2015)	7	0 Tags		247.7 MB	~5 hours ago	~5 hours
Unnamed history		0 Tags		0 bytes	Jun 27, 2016	Jun 27, 2016

Galaxy initiation - workflow

- Switch
- View
- Share or Publish
- Copy
- Rename
- Delete
- Delete Permanently

History

search datasets

Galaxy initiation - workflow

4 shown, 3 hidden

2.92 MB

7: Top exon genetic location

6: Top exons

2: Exons

1: Repeats

https://usegalaxy.org/history/list?f-sharing=All&sort=-update_time&f-name=All&f-tags=All&f-deleted=False&operation=Share+or+Publish&id=99569b6f012ffc3c

Share

Workflow



Your workflows

Create new workflow

Upload or import workflow

Name	# of Steps
Find exons with highest number of SNPs	7
Convert to tab (imported from ...)	2
imported: ChIP-seq workflow	3

- Edit
- Run
- Share or Download**
- Copy
- Rename
- View
- Delete

Workflows shared by others

No workflows have been shared by others

Other options

Configure your workflow menu

Share

Mode

The screenshot shows the Galaxy web interface for a workflow titled "Workflow ' Find exons with highest number of SNPs' ". The top navigation bar includes "Galaxy", "Analyze Data", "Workflow", "Shared Data", "Visualization", "Help", "User", and a "Using 0%" indicator. Below the navigation bar, there is a link "Go back to Workflows List". The main content area is titled "Share" and contains the following text: "This workflow is currently restricted so that only you and the users listed below can access it." There are three buttons with red arrows pointing to explanatory text: "Make Workflow Accessible via Link" points to "Restricted community"; "Make Workflow Accessible and Publish" points to "All the Galaxy server users"; "Share with a user" points to "Designated community (login@sb-roscoff.fr)". Below the sharing options, there is a section titled "Export" with a "Download" button and the text "workflow as a file so that it can be saved or imported into another Galaxy server." Below that, it says "This workflow must be accessible. Please use the option above to 'Make Workflow Accessible and Publish' before receiving a URL for importing to another Galaxy." There is also a "Create image" button and the text "of workflow in SVG format". At the bottom, it says "Export to the www.myexperiment.org site."

Share

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

→ **Restricted community**

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

→ **All the Galaxy server users**

Makes the workflow accessible via link (see above) and publishes the workflow to Galaxy's [Published Workflows](#) section, where it is publicly listed and searchable.

You have not shared this workflow with any users yet.

→ **Designated community
(login@sb-roscoff.fr)**

Export

workflow as a file so that it can be saved or imported into another Galaxy server.

This workflow must be accessible. Please use the option above to "Make Workflow Accessible and Publish" before receiving a URL for importing to another Galaxy.

of workflow in SVG format

Export to the www.myexperiment.org site.

- Get shared histories

Galaxy / METABO Analyze Data Workflow Shared Data Visualization Help User Using 216.1 MB

Tools search tools

Get Data

WORKFLOW 4 METABO...
2-Preprocessing
3-Normalisation
4-Quality Control
5-Statistical Analysis

Individual

Histories shared with you by others

Name	Datasets	Created	Last Updated	Shared by
mmonsoor	6	Apr 28, 2014	~2 days ago	mmonsoor@sb-roscoff.fr

For 0 selected histories: Copy Unshare

History

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

Galaxy / METABO Analyze Data Workflow Shared Data Visualization Help User Using 70.9 MB

Published Histories

search name, annotation, owner, and tags Advanced Search

Public

Name	Annotation	Owner	Community Tags	Last Updated
Preprocessing		mlandi		~14 seconds ago
TP1_xcms_sacuri		mmonsoor	★★★★★	~1 day ago
TP1_xcms_sacuri		jfmartin	★★★★★	Apr 28, 2014

Shared Data

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

- Get shared workflows

Your workflows

Create new workflow Upload or import workflow

Name	# of Steps
complete_workflow_RFMF	17

Individual

Workflows shared with you by others

Name	Owner	# of Steps
Workflow mmonsoor	mmonsoor@sb-roscoff.fr	7

Published Workflows

search name, annotation, owner, and tags

Public

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

Name	Annotation	Owner	Rating	Community Tags	Last Updated!
complete_workflow_RFMF		mmand			~17 hours ago

- Import shared

Galaxy / METABO Analyze Data Workflow Shared Data Visualization Help User Using 216.1 MB

Published Histories | [mmonsoor](#) | TP1 xcms sacuri **Import history** About this History

TP1 xcms sacuri

65.4 MB

search datasets

Dataset	Annotation
1: xset.RData	
2: sampleMetadata.tsv	
3: xset.TICs_raw.pdf	
4: xset.log.txt	

Author
mmonsoor

Related Histories
[All published histories](#)
[Published histories by mmonsoor](#)

Rating
Community (0 ratings, 0.0 average)
Yours

Tags
Community: none

Histories

Galaxy / METABO Analyze Data Workflow Shared Data Visualization Help User Using 216.1 MB

Your workflows

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

Name	# of Steps
complete_workflow_RFMF	17

Workflows shared with you by others

Name	Owner	# of Steps
Workflow mmonsoor	mmonsoor@sb-roscoff.fr	7

Other

- View
- Run
- Copy**
- Remove



Level 5

- Share tools and descriptions in the ToolShed

Level 4



- Launch tools autonomously
- Use advanced parameters
- Use the Galaxy API
- Provide workflow for colleagues Level 1-3

Level 3



- Launch tools autonomously
- Use workflow more or less preset

Level 2



- Use preset workflow

Level 1



- Share his data to colleagues Level 2-5

Hands-on
WORKFLOW





Which coding exon has the highest number of ~~single nucleotide polymorphisms~~ on chromosome 22?
repeats

1. Extract a workflow from your history
2. Edit the workflow (hide intermediate steps, rename inputs/outputs, set parameters at runtime, save)
3. Create a new history with the input data
4. Run the workflow
5. Share your history/workflow with your neighbour

CONCLUSION

- Easy-to-use graphical user interface for often complex command-line tools
- Keeps a full record of your analysis in a history
- Workflows enable you to repeat your analysis on different data
- Galaxy can connect to external sources for data import and visualization purposes
- Galaxy provides ways to share your results and methods with others

125+ platforms for using Galaxy: <https://galaxyproject.org/use>
[Servers, clouds, deployable resources]

GENERAL PURPOSE:

Use Galaxy servers:

usegalaxy.org, usegalaxy.eu, usegalaxy.org.au, usegalaxy.fr (coming)

DOMAIN SPECIFIC:



RNAseq: <http://galaxy.sb-roscoff.fr>

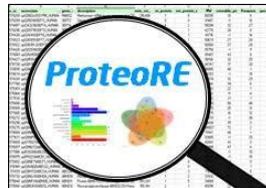
SBR tools: <http://webtools.sb-roscoff.fr>

Metagenomics: <http://galaxy4metab.sb-roscoff.fr>

Metabolomics:



Proteomics:



ChIP-seq:



Catalog of French Galaxy tools: <http://galaxycat.france-bioinformatique.fr>

- On your own:



- Training materials:

- <https://galaxyproject.github.io/training-material>

- Interactive tours of Galaxy:

- <http://galaxy.sb-roscoff.fr/tours>

- Training courses:

	Training	What ?	Where ?	When ?
	Galaxy Community Conference (GCC)	General purpose (data-intensive biology and Galaxy)	Freiburg, Germany	July 2019
	RNAseq analysis with Galaxy	RNAseq	Roscoff, France	Autumn 2019
	Workflow4Experimenters	Metabolomics	Brussels, Belgium	February 2020

END

Thank you for completing the training evaluation questionnaire:

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

BONUS

How are tools born?

BONUS

- How to import a tool in Galaxy?

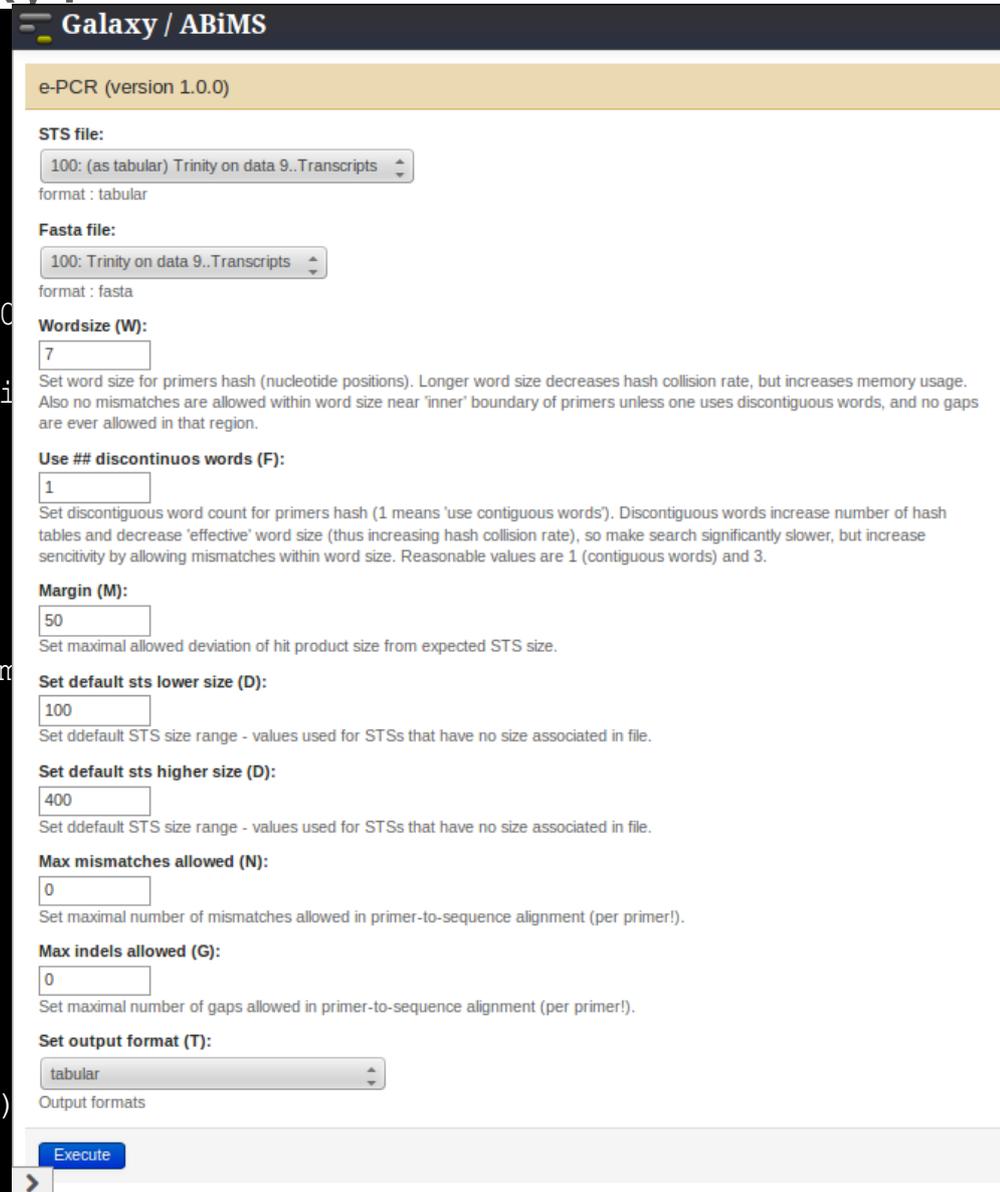


• How to import a tool in Galaxy?

```

[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 ## discontinuos words, slow i
                ##>1
  -o ##      Set output file
  -t ##      Set output format:
                1 - classic, range (pos1..pos2)
                2 - classic, midpoint
                3 - tabular
                4 - tabular with alignment in com
                    (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 - Allways 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 ## discontinuos 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

• How to import a tool in 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 ## discontinuos 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 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!)." />
>     <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>

```

• How to import a tool in 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 i
                ##>1
    -o ##      Set output file
```

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):

```
<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 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
```