



Galaxy

**An open, web-based platform for
data intensive research.**



Galaxy



- ✧ Galaxy is a framework for integrating computational tools. More @ <http://galaxy.psu.edu/>
- ✧ It allows nearly any tool that can be run from the command line to be wrapped in a well-defined interface.
- ✧ On top of these tools, Galaxy provides an accessible environment for interactive analysis that transparently tracks the details of analysis steps, a workflow system for convenient reuse, data management, sharing, publishing, and so on.
- ✧ Two (free) ways to perform analysis with Galaxy:
 - Using the open web server @ <http://main.g2.bx.psu.edu/>
 - Install his own instance - Tuto @ [http://wiki.g2.bx.psu.edu/Admin/Get Galaxy/](http://wiki.g2.bx.psu.edu/Admin/Get_Galaxy/)
- ✧ Another way is to use the Amazon Galaxy cloud

Galaxy homepage

menu

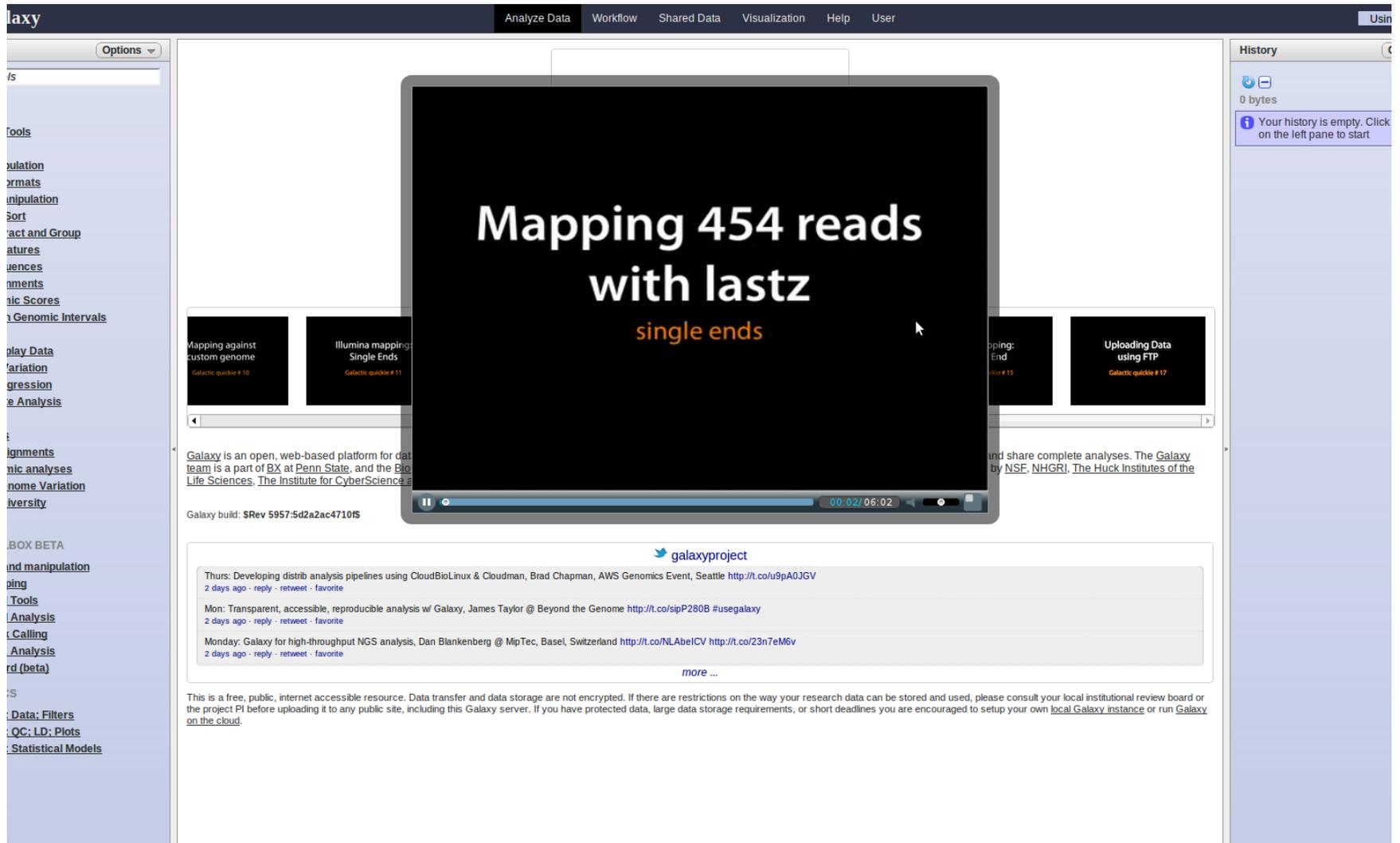
The screenshot shows the Galaxy homepage with the following components:

- Navigation Menu:** Analyse Data, Workflow, Shared Data, Visualization, Help, User
- Tools Sidebar:** search tools, Get Data, Send Data, ENCODE Tools, Lift-Over, Text Manipulation, Convert Formats, FASTA manipulation, Filter and Sort, Join, Subtract and Group, Extract Features, Fetch Sequences, Fetch Alignments, Get Genomic Scores, Operate on Genomic Intervals, Statistics, Graph/Display Data, Regional Variation, Multiple regression, Multivariate Analysis, Evolution, Motif Tools, Multiple Alignments, Metagenomic analyses, Human Genome Variation, Genome Diversity, EMBOSS, NGS TOOLBOX BETA, NGS: QC and manipulation, NGS: Mapping, NGS: SAM Tools, NGS: Indel Analysis, NGS: Peak Calling, NGS: RNA Analysis, NGS: Picard (beta), RGENETICS, SNPWGA: Data; Filters, SNPWGA: QC; LD; Plots, SNPWGA: Statistical Models
- Principal Page:** Galaxy 101 Start small The very first tutorial you need. Live Quickies: Mapping against custom genome, Illumina mapping: Single Ends, Illumina mapping: Paired Ends, Basic fastQ manipulation, Advanced fastQ manipulation, 454 Mapping: Single End, Uploading Data using FTP.
- History Sidebar:** 0 bytes, Your history is empty. Click 'Get Data' on the left pane to start.

tools

Principal page

history



The screenshot shows the Galaxy web interface. At the top, there is a navigation bar with tabs for 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. On the left side, there is a sidebar with a search bar and a list of tools categorized by function, such as 'Genomic Intervals', 'Variation', and 'Analysis'. The main content area features a video player with a large black overlay containing the text 'Mapping 454 reads with lastz single ends'. Below the video player, there is a social media feed for 'galaxyproject' with several tweets. At the bottom of the main content area, there is a disclaimer: 'This is a free, public, internet accessible resource. Data transfer and data storage are not encrypted. If there are restrictions on the way your research data can be stored and used, please consult your local institutional review board or the project PI before uploading it to any public site, including this Galaxy server. If you have protected data, large data storage requirements, or short deadlines you are encouraged to setup your own local Galaxy instance or run Galaxy on the cloud.'

Galaxy

- **NGS (Illumina, Roche-454, AB-SOLiD)**
 - QC manipulation
 - mapping (BWA/Bowtie)
 - SAM tools
 - Indel analysis
 - RNA analysis
 - ...
- **SNP**
 - Filters (Varscan)
 - QC, LD plots
 - statistical Models
- **General treatment data**
 - Get Data: upload file, BioMart, DB server, ...
 - Text manipulation: add, merge, cut columns, ...
 - Fasta manipulation: convert, compute fasta files, ...
 - Statistic: summary, count, correlation, ...
 - Graph
 - Alignments
 - Evolution
 - Metagenomics
 - ...

Existing Tools

- **FASTX-Toolkit**: tools for FASTA / FASTQ files preprocessing.
- **BWA / Bowtie**: mapping softwares, particularly suitable for short reads alignment (in paired-ends or single-ends) against one reference genome (Burrows – Wheeler Alignment tool).
- **SAMtools**: toolkit for working on the output SAM file (BWA, Bowtie, ...).
- **VarScan**: software used to filter SNPs and small indels by:
 - coverage
 - number of variant
 - base quality
 - variant allele frequency
 - pValue

How to upload your data ?

a. Tools → Get Data → Upload File

Tools

- Get Data
 - Upload File from your computer

Upload File

File Format:
 Auto-detect (dropdown menu)
 Which format? See help below

File:
 Parcourir... (button)

URL/Text:
 [Text area]

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

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

Genome:
 Click to Search or Select Build

Execute (button) **Help** (link)

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

Ab1
 A binary sequence file in 'ab1' format with a '.ab1' file extension. You must manually select this 'File Format' when uploading the file.

Axt
 blastz pairwise alignment format. Each alignment block in an axt file contains three lines: a summary line and 2 sequence lines. Blocks are separated from one another by blank lines. The summary line contains chromosomal position and size information about the alignment. It consists of 9 required fields.

Bam
 A binary file compressed in the BGZF format with a '.bam' file extension.

How to upload your data ?

The screenshot shows the Galaxy web interface with a navigation menu on the left and a main workspace. A red box highlights a notification message: "Your upload has been queued. History entries that are still uploading will be blue, and turn green upon completion. Please do not use your browser's 'stop' or 'reload' buttons until the upload is complete, or it may be interrupted. You may safely continue to use Galaxy while the upload is in progress. Using 'stop' and 'reload' on pages other than Galaxy is also safe." The history panel on the right shows a file named "2: Genome_Tomate.fasta" which is highlighted in blue, indicating it is still uploading.

The legend below the notification explains the status of history entries:

- Waiting**: Represented by a grey entry "5: (BWA) Output SAM".
- Ongoing**: Represented by a blue entry "2: Genome_Tomate.fasta".
- Finished → Error**: Represented by a red entry "5: (BWA) Output SAM".
- Finished → OK**: Represented by a green entry "3: Genome_Tomate.fasta".

How to upload your data ?

Example

The screenshot shows the Galaxy web interface with the 'Upload File (version 1.1.3)' tool selected. The tool's options include 'File Format' (set to 'Auto-detect'), 'File' (with a 'Parcourir...' button), 'URL/Text' (with a text area), 'Files uploaded via FTP' (with a table), 'Convert spaces to tabs' (with a 'Yes' checkbox), and 'Genome' (with a dropdown menu). An 'Execute' button is at the bottom.

Overlaid on the interface is a file selection dialog titled 'Envoi du fichier'. The dialog shows a list of files in the 'ylo' directory. The file 'inputCR.gff3' is selected.

Nom	Taille	Modifié
inputmapping.map	870 octets	17/06/2011
inputFileTest2.bed	147 octets	13/12/2011
inputFileTest1.bed	282 octets	13/12/2011
inputCR.gff3	956 octets	13/12/2011
GetWigProfile_transcriptFile.gff3	47 octets	15/12/2011
GetWigPro_file.wig	125 octets	15/12/2011
getRD.fasta	42 octets	20/12/2011
GetDiffExpr_refFile.gff3	52 octets	15/12/2011
GetDiffExpr_inputFile2.gff3	135 octets	15/12/2011
GetDiffExpr_inputFile1.gff3	135 octets	15/12/2011
expRef.fasta	1,9 Mo	17/06/2011
expOutputGff.gff3	464,6 Ko	13/02/2012
clusterize_strands_expected.gff3	309,3 Ko	13/02/2012
clusterize_output_tag_expected.gff3	169,3 Ko	13/02/2012
clusterize_normalize_expected.gff3	153,6 Ko	13/02/2012
clusterize_default_expected.gff3	153,6 Ko	13/02/2012
adress.txt	25 octets	17/06/2011

How to upload your data ?

Example

Galaxy

Analyze Data Workflow Shared Data Visualization Help User Using 0%

Tools Options

search tools

Get Data

- Upload File from your computer
- UCSC Main table browser
- UCSC Archaea table browser
- BX main browser
- EBI SRA ENA SRA
- BioMart Central server
- GrameneMart Central server
- Flymine server
- modENCODE fly server
- modENCODE modMine server
- Ratmine server
- YeastMine server
- modENCODE worm server

chr1	test	match	6155418	6155441	24	+	.	Name=test1/1;occurrence=1;rank=1;bestRegion=(self);nt
chr2	test	match	26303950	26303981	32	+	.	Name=test2/1;occurrence=1;rank=1;best
chr3	test	match	28320540	28320574	35	+	.	Name=test2/1;occurrence=2;bestRegion=
chr4	test	match	28565007	28565041	35	+	.	Name=test2/1;occurrence=3;rank=3;best
chr1	test	match	6155418	6155441	24	+	.	Name=test3/1;occurrence=2;rank=2;bestRegion=(self);nt
chr1	test	match	6155418	6155441	24	-	.	Name=test3/1;occurrence=2;rank=2;bestRegion=(self);nt

History Options

0 bytes

You are currently viewing a deleted history!

1: inputCR.gff3

6 lines

format: gff3, data

Info: uploaded gff3 file

1. Seqid	2. Source	3. Type	4. Start	5. End
chr1	test	match	6155418	61554
chr2	test	match	26303950	26303
chr3	test	match	28320540	28320
chr4	test	match	28565007	28565
chr1	test	match	6155418	61554
chr1	test	match	6155418	61554

How to use a tool ?

a. Choose a tool from the list Example

The screenshot shows the Galaxy web interface with the 'Cut' tool selected. The tool configuration panel is highlighted with a red box around the 'Execute' button. A red arrow points from the 'Execute' button to the 'Cut' tool title. A green arrow points from the 'Cut columns' field to the 'Delimited by' field. An orange arrow points from the 'Execute' button to the 'Execute' button. A red arrow points from the 'Execute' button to the 'Execute' button.

Tools (Options ▾)

Text Manipulation

- Add column to an existing dataset
- Compute an expression on every row
- Concatenate datasets tail-to-head
- Condense consecutive characters
- Convert delimiters to TAB
- Merge Columns together
- Create single interval as a new dataset
- Cut columns from a table**
- Change Case of selected columns
- Paste two files side by side
- Remove beginning of a file
- Select random lines from a file
- Select first lines from a dataset
- Select last lines from a dataset
- Trim leading or trailing characters
- Line/Word/Character count of a dataset
- Secure Hash / Message Digest on a dataset

Convert Formats

FASTA manipulation

Filter and Sort

Join, Subtract and Group

Extract Features

Fetch Sequences

Fetch Alignments

Get Genomic Scores

Operate on Genomic Intervals

Statistics

Graph/Display Data

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

Cut (version 1.0.1)

Cut columns:
c1, c4, c5,c7

Delimited by:
Tab

From:
1: inputCR.gff3

Execute

b. Set options and parameters

c. Execute

WARNING: This tool breaks column assignments. To re-establish column assignments run the tools and click on the pencil icon in the latest history item.

The output of this tool is always in tabular format (e.g., if your original delimiters are commas, they will be replaced with tabs). For example:

```

Cutting columns 1 and 3 from:

apple,is,good
windows,is,bad

will give:

apple good
windows bad
    
```

What it does

This tool selects (cuts out) specified columns from the dataset.

Columns are specified as **c1**, **c2**, and so on. Column count begins with **1**

Columns can be specified in any order (e.g., **c2,c1,c6**)

if you specify more columns than actually present - empty spaces will be filled with dots

Example

input dataset (six columns: c1, c2, c3, c4, c5, and c6):

```

chr1 10 1000 gene1 0 +
chr2 100 1500 gene2 0 +
    
```

cut on columns "c1,c4,c6" will return:

```

chr1 gene1 +
chr2 gene2 +
    
```

cut on columns "c6,c5,c4,c1" will return:

```

0 gene1 chr1
0.gene2.chr2
    
```

History (Options ▾)

1.9 Kb

1: inputCR.gff3 (eye icon) (edit icon) (delete icon)

6 lines
format: gff3, database: ?
Info: uploaded gff3 file

1. Seqid	2. Source	3. Type	4. Start	5. End
chr1	test	match	6155418	61554
chr2	test	match	26303950	26303
chr3	test	match	28320540	28320
chr4	test	match	28565007	28565
chr1	test	match	6155418	61554
chr1	test	match	6155418	61554

Help

Display results

Example

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

Tools Options ▾

Text Manipulation

- [Add column](#) to an existing dataset
- [Compute](#) an expression on every row
- [Concatenate datasets](#) tail-to-head
- [Condense](#) consecutive characters
- [Convert](#) delimiters to TAB
- [Merge Columns](#) together
- [Create single interval](#) as a new dataset
- [Cut](#) columns from a table
- [Change Case](#) of selected columns
- [Paste](#) two files side by side
- [Remove beginning](#) of a file
- [Select random lines](#) from a file
- [Select first lines](#) from a dataset
- [Select last](#) lines from a dataset
- [Trim](#) leading or trailing characters
- [Line/Word/Character count](#) of a dataset
- [Secure Hash / Message Digest](#) on a dataset

chr1	6155418	6155441	+
chr2	26303950	26303981	+
chr3	28320540	28320574	+
chr4	28565007	28565041	+
chr1	6155418	6155441	+
chr1	6155418	6155441	-

History Options ▾

1.9 Kb

 You are currently viewing a deleted history!

6: Cut on data 1 👁 ✎ ✕

6 lines
format: tabular, da

[Display data in browser](#)

1	2	3	4
chr1	6155418	6155441	+
chr2	26303950	26303981	+
chr3	28320540	28320574	+
chr4	28565007	28565041	+
chr1	6155418	6155441	+
chr1	6155418	6155441	-

1: inputCR.gff3 👁 ✎ ✕

Display results

Example

The screenshot displays the Galaxy web interface. The central panel shows a volcano plot of differential expression analysis results. The x-axis is labeled 'mean' and ranges from 10^1 to 10^5 . The y-axis is labeled 'log₂ fold change' and ranges from -5 to 5. A dense cloud of black points is centered around the origin. Red points, representing differentially expressed genes, are scattered around the periphery. A 'History' panel on the right lists various files, including '34: [DiffExpAna] Output PNG File', which has a tooltip that says 'Display data in browser'.

Jpg, png, gif, ...

Download results

Example

```

esVarA resVarB
76508064 1.05468295412678e-07 1.26823515867838e-05 0.000934278538040018 0.020730024317302
.0403950249907778 0.741913880409664 0.976848943275899 0.0142093862358696 0.676502120827453
.0711320287863404 0.61389189534877 0.936602626775006 0.660288183718132 0.774466743898237
3.400585488220568 0.00801890500699104 0.124645461385814 0.5310547981488 0.335385991358929
3.658507228674406 0.00508796945912739 0.092252435392212 0.909915797374303 0.689563019140147
3.574620349388894 0.355017583390741 0.77382095358849 0.538411276444988 0.633481921460258

.851966611221824 0.456163534474002 0
2.80987138393820 0.111878269277407 0.512383383762557 0.38669328071756 0.335431060567449
3497137013 0.97631599469946 1 9.87497015303255e-06 1.37776993327580 0.335431060567449
3.247914463467403 0.0530509603617875 0.363475757704318 0.0305839005578333 1.17896077677876

.635575738799353 1 1 0.654604367316726 1.33634292360821 2.02012728179680
.313622892210804 0.498131508398385 0.85744832703999 0.120137110650941 2.02012728179680

495352674 0.482438636458077 0.853450290615918 1.44796119570087 0.004550928209665

34796926 0.0167290013673535 0.192906878283985 0.843371953189259 4.28053779180959
3.255433181545583 0.0342576062415289 0.286418563207185 0.407125506635714 4.35622505883254
4192051 6.92880408168522e-10 1.23984350180727e-07 0.168055577386821 5.54692023792936
.171512467355832 0.179215948849292 0.623852461128697 0.00611123091303218 0.626503998946387
.168397159101002 0.228099055330641 0.679922628785033 0.00870698196068093 0.441020834366558
.360976545289841 0.0243184735725121 0.234571061759830 0.358052642674222 5.82363888832967
  
```

DOWNLOAD TO YOUR LOCAL PC

```

3.0726865010426558 0.50531483481245 0.86281804862728 2.82537614839373 2.05292461881673
3.0133583411688436 1 1 0.111201901150546 0.265418689145608
471034147 0.475452647787548 0.851966611221824 0.98839004011543 0.0155946648136824
.176305429983453 0.908749984845632 1 0.000107335015650952 0.236943416837597
3.418942123979095 0.075861625372282 0.433235596873393 4.75763193879995 2.07980629496884
.0521308234699158 0.691338118375052 0.973354902019312 0.00182229479291425 1.25154030333775
.00154955224767693 0.984451591572817 1 0.332822922931682 2.74937606892162
3.106261766582178 0.565489115474195 0.90619906981614 0.195967517155396 1.95689714558609
370323286 0.25594093226543 0.715374328828468 0.0478570619773775 0.0738679981228459
3.384839979130366 0.818500087137336 1 0.00553628350268612 0.0963911504841196
3.384839979130366 0.818500087137336 1 0.00553628350268612 0.0963911504841196
.470591642459686 0 1.04570830289961
.464679121570021
3.13944362911238 0.731587532080332 0.973354902019312 0.157680800685101 0.0760113750164207
.197820880764164 0.3726327444785024 0.787769730923163 0.787769730923163 0.438948737513628
.092902095587149 0.666230630880457 0.9672501608171177 1.82282357285413 0.253286107905144
347980264275991 0.0295371101841186 0.262446091457794 2.705012247437 0.558505430853262
46328218 0.411634713136816 0.818094721215071 0.23673066922225 0.553910948851516
.0993606916211818 0.713937694196624 0.973354902019312 0.215308831919737 0.0198927608286350
  
```

upload files 1.1 Gb

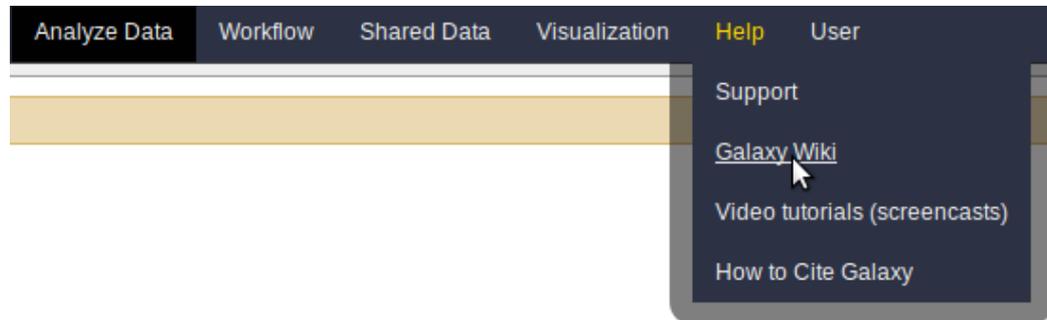
- 26: [DiffExpAna] Log File
- 25: [DiffExpAna] Output PNG File
- 24: [DiffExpAna] Output CSV File
 - 17,606 lines
 - format: tabular
 - Info: locfit 1.5-6 2010-01-20
 - [1] 4
 - null device
- 14: fly RNA counts.tsv
- 8: NC 011753 annot

Download icon highlighted with a red box and arrow.

1	2	3	4
id	baseMean	baseMeanA	
1 Gene_1	14.1198573935869	26.74501637693	
2 Gene_2	12190.3236903247	12019.67215019	
3 Gene_3	605.824158472026	590.8921137818	
4 Gene_4	993.086721892094	1130.080257943	
5 Gene_5	183.515345279972	224.6851348364	

Where can I get help?

- **Galaxy Wiki:**
<http://wiki.g2.bx.psu.edu/>



- **Galaxy mailing lists:**
<http://wiki.g2.bx.psu.edu/Mailing%20Lists>



URGI Galaxy

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

Tools Options

search tools

Get Data
Send Data
ENCODE Tools
Lift-Over
Text Manipulation
Convert Formats
FASTA manipulation
Filter and Sort
Join, Subtract and Group
Extract Features
Fetch Sequences
Fetch Alignments
Get Genomic Scores
Operate on Genomic Intervals
Statistics
Graph/Display Data
Regional Variation
Multiple regression
Multivariate Analysis
Evolution
Motif Tools
Multiple Alignments
Metagenomic analyses
Human Genome Variation
Genome Diversity
EMBOSS

NGS TOOLBOX BETA
NGS: QC and manipulation
NGS: Mapping
NGS: SAM Tools
NGS: Indel Analysis
NGS: Peak Calling
NGS: RNA Analysis
<http://urgi.versailles.inra.fr/galaxy>

Galaxy @ URGI
Analyze mapped RNA-Seq data, discover SNPs, ...

Live Quickies

- Illumina mapping: Paired Ends (Galactic quickie # 12)
- Basic fastQ manipulation: (Galactic quickie # 13)
- Advanced fastQ manipulation: (Galactic quickie # 14)
- 454 Mapping: Single End (Galactic quickie # 15)
- Uploading Data using FTP (Galactic quickie # 17)

Galaxy is an open, web-based platform for data intensive biomedical research. Whether on this free public server or your own instance, you can perform, reproduce, and share complete analyses. 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 NSF, NHGRI, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Emory University.

Galaxy build: **SRev 6801:40f1816d16857S**

galaxyproject

- galaxyproject usegalaxy.org is down until ~ 5:45PM EDT (UTC -0400). Jobs will not be interrupted during this period. Rebuilding database tables... 19 hours ago · reply · retweet · favorite
- galaxyproject usegalaxy.org down briefly today as we move & upgrade system services. Should be < 15 minutes and will not interrupt running jobs. yesterday · reply · retweet · favorite
- galaxyproject @greg_vonkuster: Speed improvements should be noticed in the Galaxy tool shed! #usegalaxy bit.ly/gxyshed yesterday · reply · retweet · favorite

[more ...](#)

History Options

0 bytes

You are currently viewing a deleted history!

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

Tools integration in URGI Galaxy

<http://urgi.versailles.inra.fr/>



The screenshot shows the URGI website interface. At the top, there is a navigation bar with links for FEEDBACK, CONTACT, SITE MAP, and ABOUT US, along with a Register button. The main header features the URGI logo and the text "PLANT AND FUNGI DATA INTEGRATION". Below this is a green navigation bar with tabs for Platform, Research, Projects, Data, Tools, and Species. The Tools tab is selected, and a red box highlights a circular menu of tools including GnpMap, GnpSeq, GnpSnp, GnpGenome, SReGal, Ephesis, GnpArray, and GnpProt. A red arrow points from the GnpGenome tool to the "GNPIS PORTAL" section below. The "SPECIES" section lists various tools like structural variants, genomics, Pipelines, REPET package, TE annotation, heterochromatin, molecular evolution, TE classification, and Transposable Elements. The "RESEARCH" section is also visible. On the right side, there is a "WHAT'S NEW?" section with an RSS icon, listing updates for GnpIS, GnpMap, GnpSNP, and Siregal, each with a gear icon and a link to a private data server. At the bottom, there is an "EVENTS" section with an RSS icon, listing events such as "Transposable Elements in Marine Stramenopiles" and "Vitis vinifera annotation jamboree".

Tools integration in URGI Galaxy

The screenshot displays the URGI Galaxy website interface. At the top, the URGI logo and the text "GnplS GENETIC AND GENOMIC INFORMATION SYSTEM" are visible, along with a "FEEDBACK" link. The main content area is divided into several sections:

- Searches:** Includes "QUICK SEARCH" and "ADVANCED TOOLS".
- Documentation:** Includes "USER GUIDE", "NEWS", and "RELEASE NOTES".
- Data:** Includes "DATA SUBMISSION", "ARRAYS", "GENOMES", "TAXONS", "GENETIC MAPS", "SEQUENCES", "POLYMORPHISMS", "PHENOTYPES", "GENETIC RESOURCES", and "PLANT SYNTENY".

The "QUICK SEARCH" form is highlighted, showing the input "VVIF52" and a "SUBMIT" button. Below the input, it says "You can find the indexed data list [here](#)." and provides examples: "VV1, VVIF52, gene, arabidopsis, AY109603, Xwmc430".

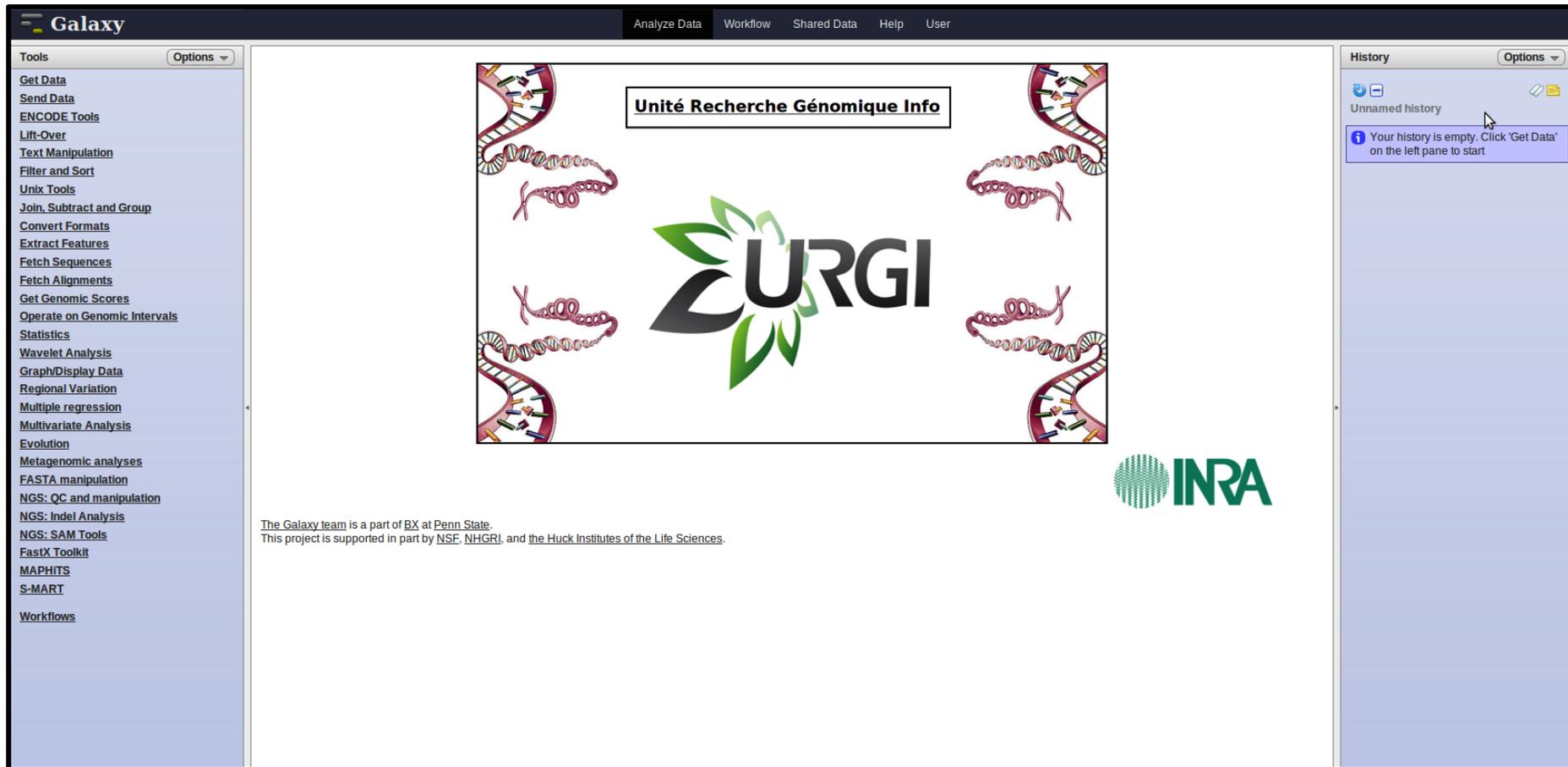
The "ADVANCED TOOLS" section is also highlighted, with the "GALAXY" button circled in red and a red arrow pointing to it. A yellow callout box next to the "GALAXY" button says "Advanced search using Galaxy".

On the right side, there are several tool categories listed:

- Genomes:** Genome annotation data. GnpGenome.
- Taxons:** Taxonomic data.
- Sequences:** NGS projects description. GnpSeq.
- Genetic maps:** Genetic maps and QTLs. GnpMap.
- Polymorphisms:** Molecular polymorphism. GnpSNP.
- Phenotypes:** Phenotypic and environmental experiments. Ephasis.
- Genetic resources:** Plant genetic resources data. Siregal.
- Arrays:** Expression data. GnpArray.

At the bottom left, the URL "ailles.inra.fr/galaxy" is visible.

Tools integration in URGI Galaxy



Galaxy

Analyze Data Workflow Shared Data Help User

Tools Options

- Get Data
- Send Data
- ENCODE Tools
- Lift-Over
- Text Manipulation
- Filter and Sort
- Unix Tools
- Join, Subtract and Group
- Convert Formats
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Wavelet Analysis
- Graph/Display Data
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Evolution
- Metagenomic analyses
- FASTA manipulation
- NGS: QC and manipulation
- NGS: Indel Analysis
- NGS: SAM Tools
- FastX Toolkit
- MAPHITS
- S-MART
- Workflows

Unité Recherche Génomique Info

URGI

INRA

History Options

Unnamed history

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

The Galaxy team is a part of BX at Penn State.
This project is supported in part by NSF, NHGRI, and the Huck Institutes of the Life Sciences.

Installation of URGI Galaxy

Galaxy is installed on URGI cluster with:

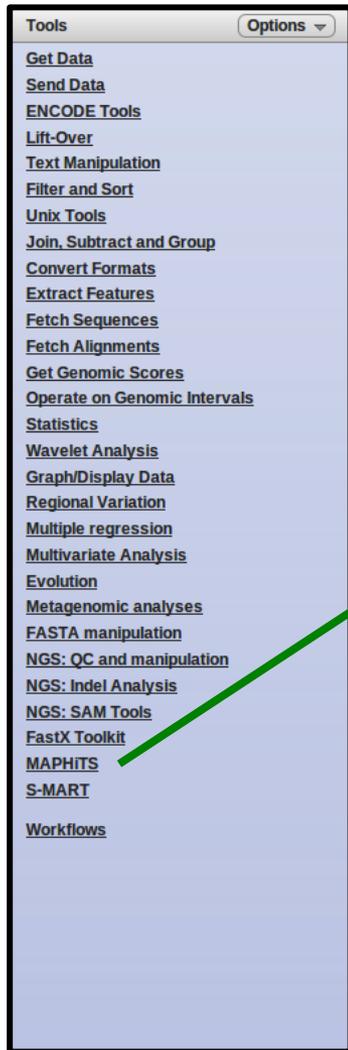
- CPU: **704** (Intel Xeon)
- RAM max: **96 Gb** per job
- Storage: **60 Tb**



Using Sun Grid Engine (for job management) and a PostgreSQL Database (for Galaxy).

<http://urgi.versailles.inra.fr/galaxy>

New URGI Integrated tools



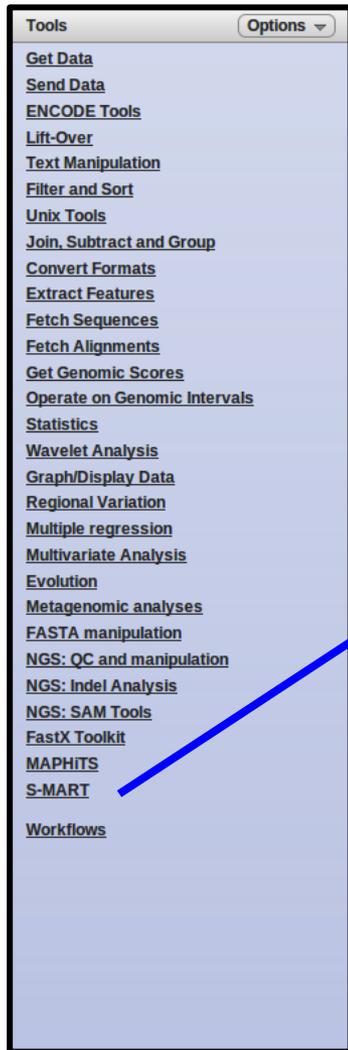
MAPHITS

MAPHITS

PREPROCESS TOOLS

- Header fasta filter Remove all informations in each header of fasta file.
- Remove duplicate short reads
- Remove duplicate short reads for big files (> 2Go)
- Remove short reads not in paired-ends
- Remove short reads not in paired-ends for big files (>2Go)
- Remove short reads > N %
- Remove short reads > N % for big files (>2Go)

New URGI Integrated tools



S-MART

S-MART

CONVERSION TOOLS

- Gff3 -> Wig Convert Gff3 File to Wig File.
- Bed -> Csv Convert Bed File to Csv File.
- Bed -> Gff2 Convert Bed File to Gff2 File.
- Bed -> Gff3 Convert Bed File to Gff3 File.
- Bed -> Sam Convert Bed File to Sam File.
- Blast (-m 8) -> Csv Convert Blast (-m 8) File to Csv File.
- Blast (-m 8) -> Gff2 Convert Blast (-m 8) File to Gff2 File.
- Blast (-m 8) -> Gff3 Convert Blast (-m 8) File to Gff3 File.

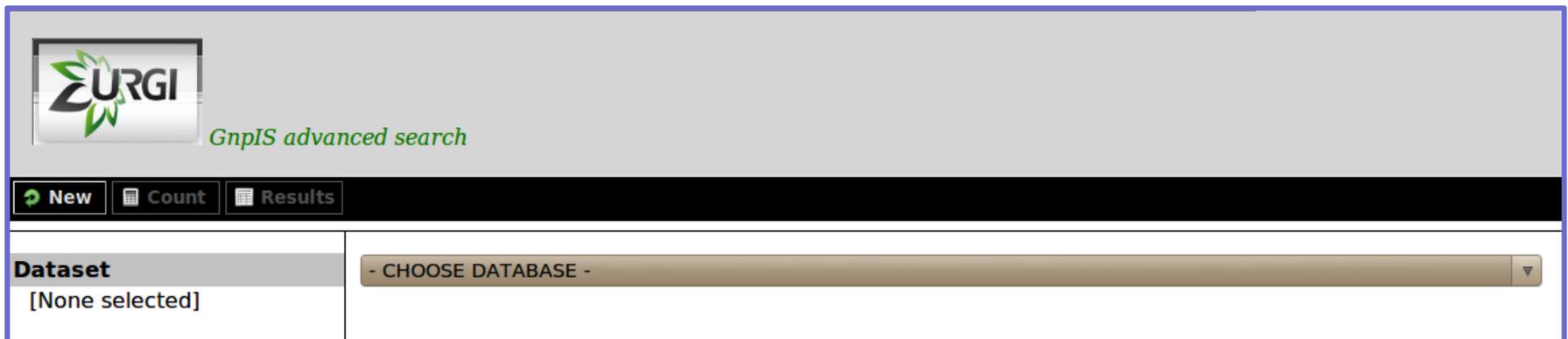
New URGI Integrated tools

Access to URGI
Information System
via **BioMart** software

Get Data:

- [Upload File](#) from your computer
- [UCSC Main](#) table browser
- [UCSC Test](#) table browser
- [UCSC Archaea](#) table browser
- [BX main](#) browser
- [Get Microbial Data](#)
- [BioMart](#) Central server
- [BioMart INRA URGI GnpIs](#)
- [CBI Rice Mart](#) rice mart
- [GrameneMart](#) Central server

BioMart
URGI
GnpIs



The screenshot shows the URGI GnpIS advanced search interface. At the top left is the URGI logo and the text "GnpIS advanced search". Below this is a navigation bar with buttons for "New", "Count", and "Results". The main area is divided into two sections: "Dataset" with the text "[None selected]" and a dropdown menu labeled "- CHOOSE DATABASE -".

Background and objectives of MAPHITS

URGI team develops a pipeline (*MAPHiTS*) for SNPs detection from short reads. The set of SNPs is between various species of Grape after mapping short reads against a reference genome.

```
-----ATGCATGCTAGCTAGACTGTACG----- (reference)
-----ATGCATACTAGCTAGACTGTACG----- (read A)
```

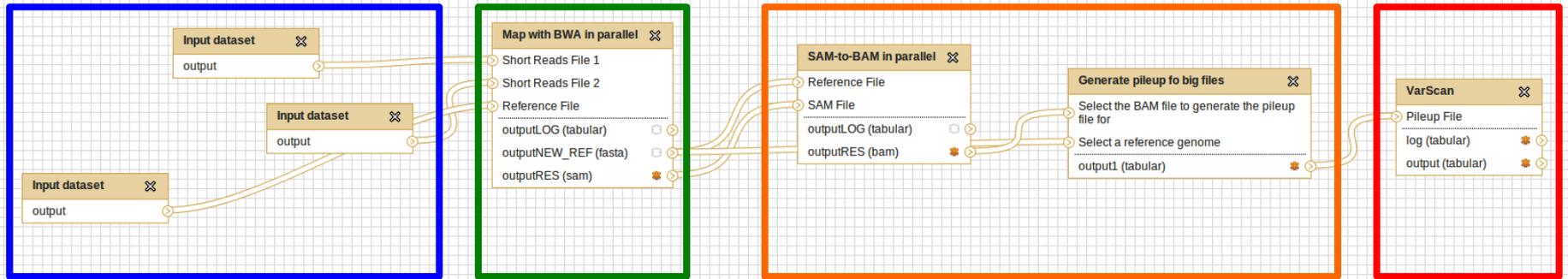


SNP

Link: <http://urgi.versailles.inra.fr/Tools/MAPHITS>

Contact: urgi-contact@versailles.inra.fr

MAPHiTS



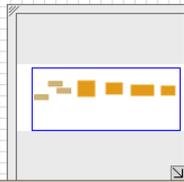
Input Files

**BWA
in
parallel**

**SAM to
BAM in
parallel**

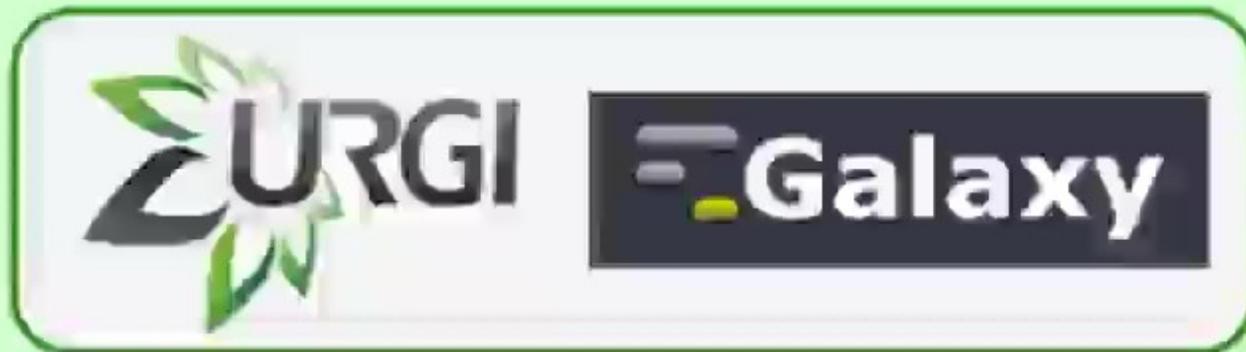
+
**Generate
Pileup**

VarScan



Use MAPHiTS in Galaxy

IX- Use MAPHiTS workflow



Background and objectives of the S-MART¹

- RNA-seq (whole transcriptome shotgun sequencing), efficient ways to measure transcriptome data experimentally.
- URGI team develops a tool (S-MART) for mapped RNA-seq data analysis. Such as mapping high-throughput sequencing data from a genome.
- Use S-MART for data manipulation, visualization, differential expression.

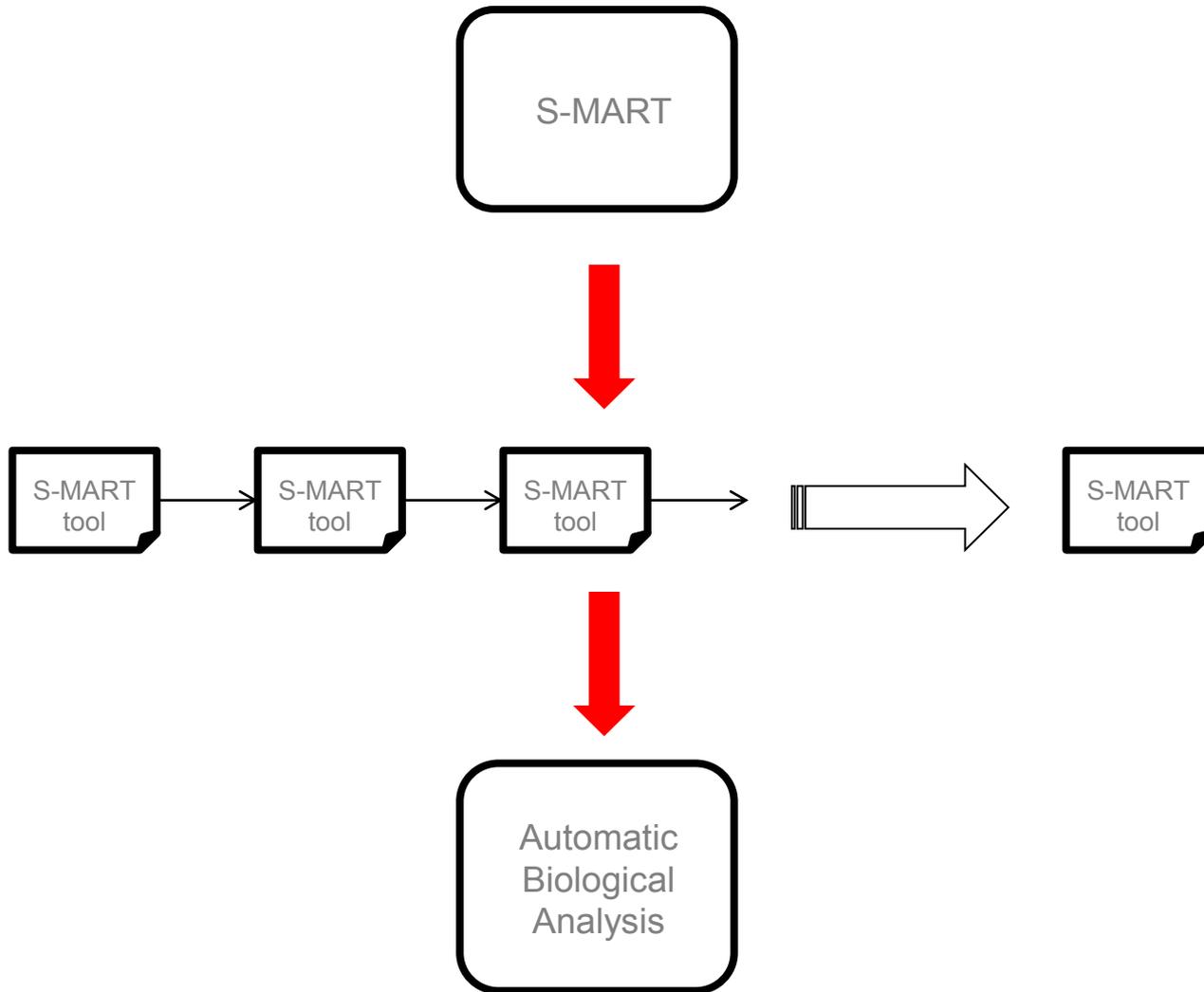
Link: <http://urgi.versailles.inra.fr/Tools/S-MART>

Contact: matthias.zytnicki@versailles.inra.fr

1: Zytnicki M, Quesneville H (2011) S-MART, A Software Toolbox to Aid RNA-seq Data Analysis. PLoS ONE 6(10):e25988.doi:10.1371/journal.pone.0025988

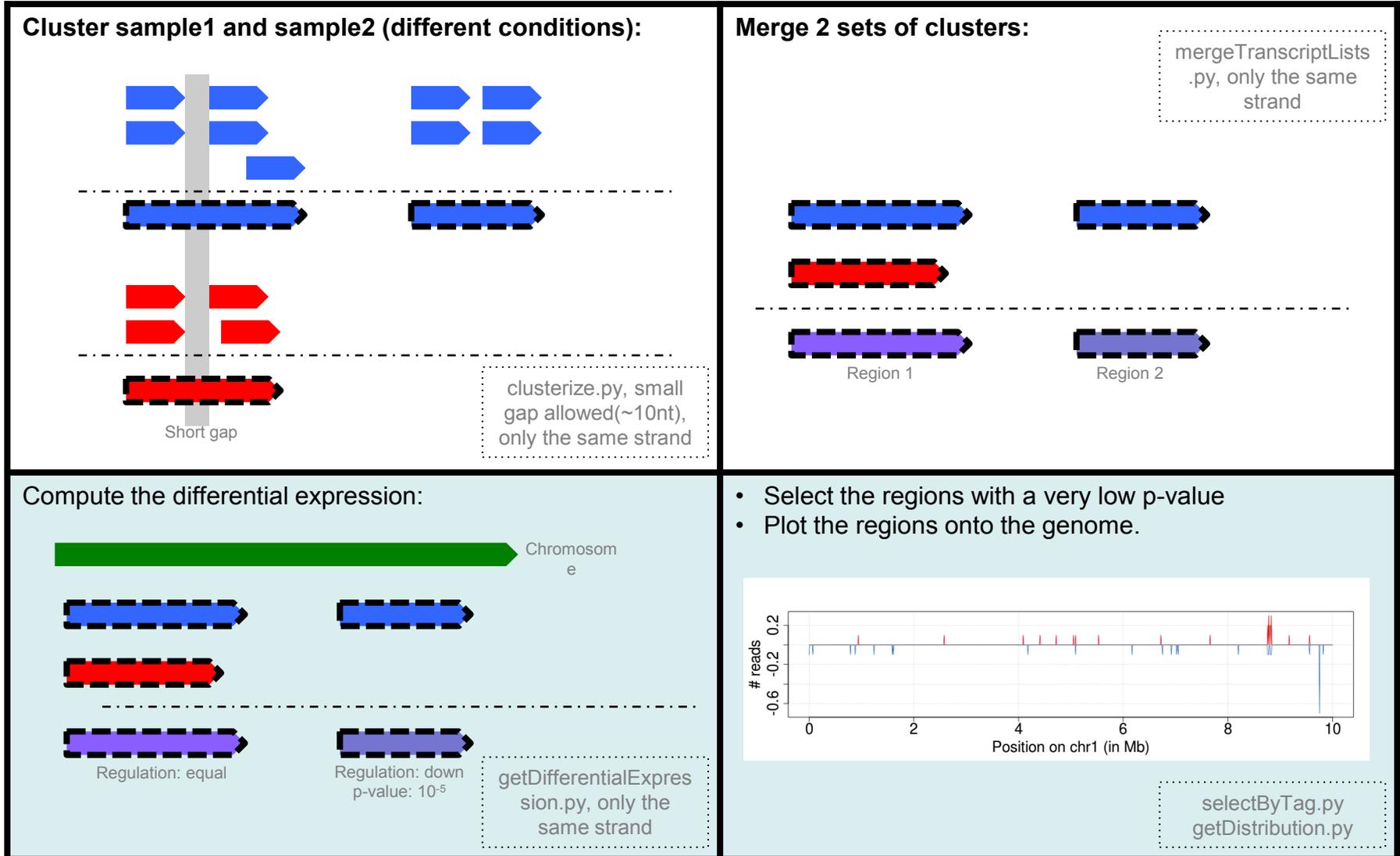
S-MART

**PIPELINE/WO
RKFLOW**



S-MART

possible pipeline: differential expression



Run S-MART workflow in Galaxy

Galaxy Analyze Data Workflow Shared Data Admin Help User

Tools Options

- Get Data
- Send Data
- ENCODE Tools
- Lift-Over
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- Convert Formats
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Wavelet Analysis
- Graph/Display Data
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Evolution
- Motif Tools
- Multiple Alignments
- Metagenomic analyses
- FASTA manipulation
- NCBI BLAST+
- NGS: QC and manipulation
- NGS: Picard (beta)
- NGS: Mapping
- NGS: Indel Analysis
- NGS: RNA Analysis
- NGS: SAM Tools
- NGS: GATK Tools (beta)
- NGS: Peak Calling
- NGS: Simulation
- SNP/WGA: Data; Filters
- SNP/WGA: QC; LD; Plots
- SNP/WGA: Statistical Models
- Human Genome Variation
- Genome Diversity
- VCF Tools
- S-MART
- PIPE-LINE
- Differential expression analysis
- Workflows

Step 1: Input dataset

.sam format

8: sample2.sam

type to filter

Step 1

Step 2: Input dataset

.sam format

8: sample2.sam

type to filter

Step 2

Step 3: Input dataset

.fasta format

9: sequence.mfa

type to filter

Step 3

Step 4: Clusterize

Input File Format
sam

Input File
Output dataset 'output' from step 1

colinear option
True

normalize option for only GFF3 file format
False

log option
False

distance option ← Define the parameter

Step 5: Clusterize

Step 6: mergeTranscriptLists

Step 7: GetDifferentialExpression

Step 8: SelectByTag

Step 9: getDistribution

Send results to a new history → Write in a new history

Run workflow

Run S-MART workflow in Galaxy

Galaxy Analyze Data Workflow Shared Data Admin Help User Using 5.8 Gb

Tools Options ▾

- Statistics
- Wavelet Analysis
- Graph/Display Data
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Evolution
- Motif Tools
- Multiple Alignments
- Metagenomic analyses
- FASTA manipulation
- NCBI BLAST+
- NGS: QC and manipulation
- NGS: Picard (beta)
- NGS: Mapping
- NGS: Indel Analysis
- NGS: RNA Analysis
- NGS: SAM Tools
- NGS: GATK Tools (beta)
- NGS: Peak Calling
- NGS: Simulation
- SNP/WGA: Data, Filters
- SNP/WGA: QC; LD; Plots
- SNP/WGA: Statistical Models
- Human Genome Variation
- Genome Diversity
- VCF Tools
- S-MART
- PIPE-LINE
- Differential expression analysis
- Workflows**
- Differential_expression
- pipeLine_ANTISENS (imported from uploaded file)
- pipeLine_CIS (imported from uploaded file)
- TRANS_new (imported from uploaded file)
- All workflows

✓ Successfully ran workflow "Differential_expression". The following datasets have been added to the queue:

These datasets will appear in a new history: 'Differential_expression'.

- 1: [clusterize]output file
- 2: [clusterize]output file
- 3: [mergeTranscriptLists]out file
- 4: [GetDifferentialExpression]out file
- 5: [SelectByTag] Output File
- 6: [getDistribution] tar out file

Submit to run the workflow

History Options ▾

upload files 2.2 Gb

12: sequence.mfa	eye	/	✕
11: sample2.sam	eye	/	✕
10: sample1.sam	eye	/	✕
9: fly_RNA_counts.tsv	eye	/	✕
8: NC 011744r known_miscRNA.gff	eye	/	✕
7: NC 011744 annot.gff	eye	/	✕
6: NC 011744 RNAseq_smart.gff3	eye	/	✕
5: NC 011753 annot.gff	eye	/	✕
4: NC 011753 RNAseq.gff3	eye	/	✕
3: NC 01174453 RNAseq.gff3	eye	/	✕
2: NC 01174453 annot.gff	eye	/	✕
1: NC 01174453r known_miscRNA.gff	eye	/	✕

Run S-MART workflow in Galaxy

on Disk	Created	Last Updated ↑	Status
	less than a minute ago	less than a minute ago	current history
	Feb 22, 2012	1 day ago	
b	4 days ago	2 days ago	
b	Feb 23, 2012	Feb 23, 2012	
	Feb 23, 2012	Feb 23, 2012	
b	Feb 23, 2012	Feb 23, 2012	
b	Feb 22, 2012	Feb 22, 2012	

History	Options
Differential_expression	0 bytes
6: [getDistribution] tar out file	View Edit Delete
5: [SelectByTag] Output File	View Edit Delete
4: [GetDifferentialExpression]out file	View Edit Delete
3: [mergeTranscriptLists]out file	View Edit Delete
2: [clusterize]output file	View Edit Delete
1: [clusterize]output file	View Edit Delete

Step 6 ←
Step 5 ←
Step 4 ←
Step 3 ←
Step 2 ←
Step 1 ←

Edit S-MART workflow in Galaxy

Galaxy

Your workflows

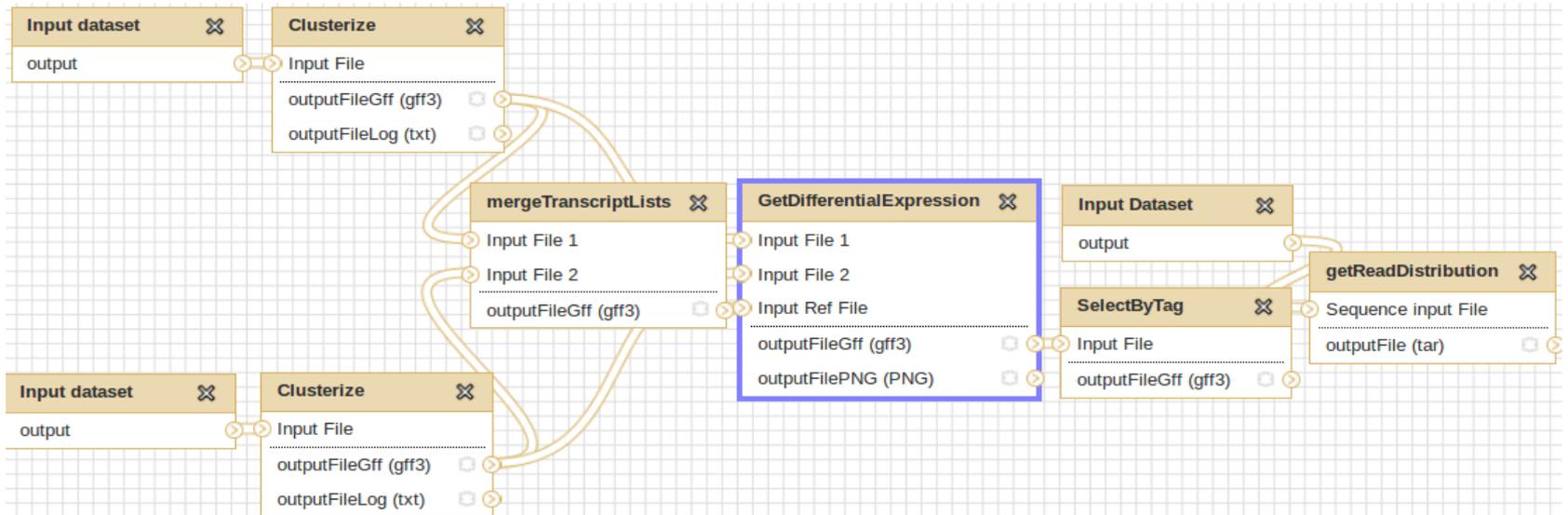
Name	Actions
Differential_expre	<ul style="list-style-type: none"> Edit Run Share or Publish Download or Export Clone Rename Delete

Workflows S

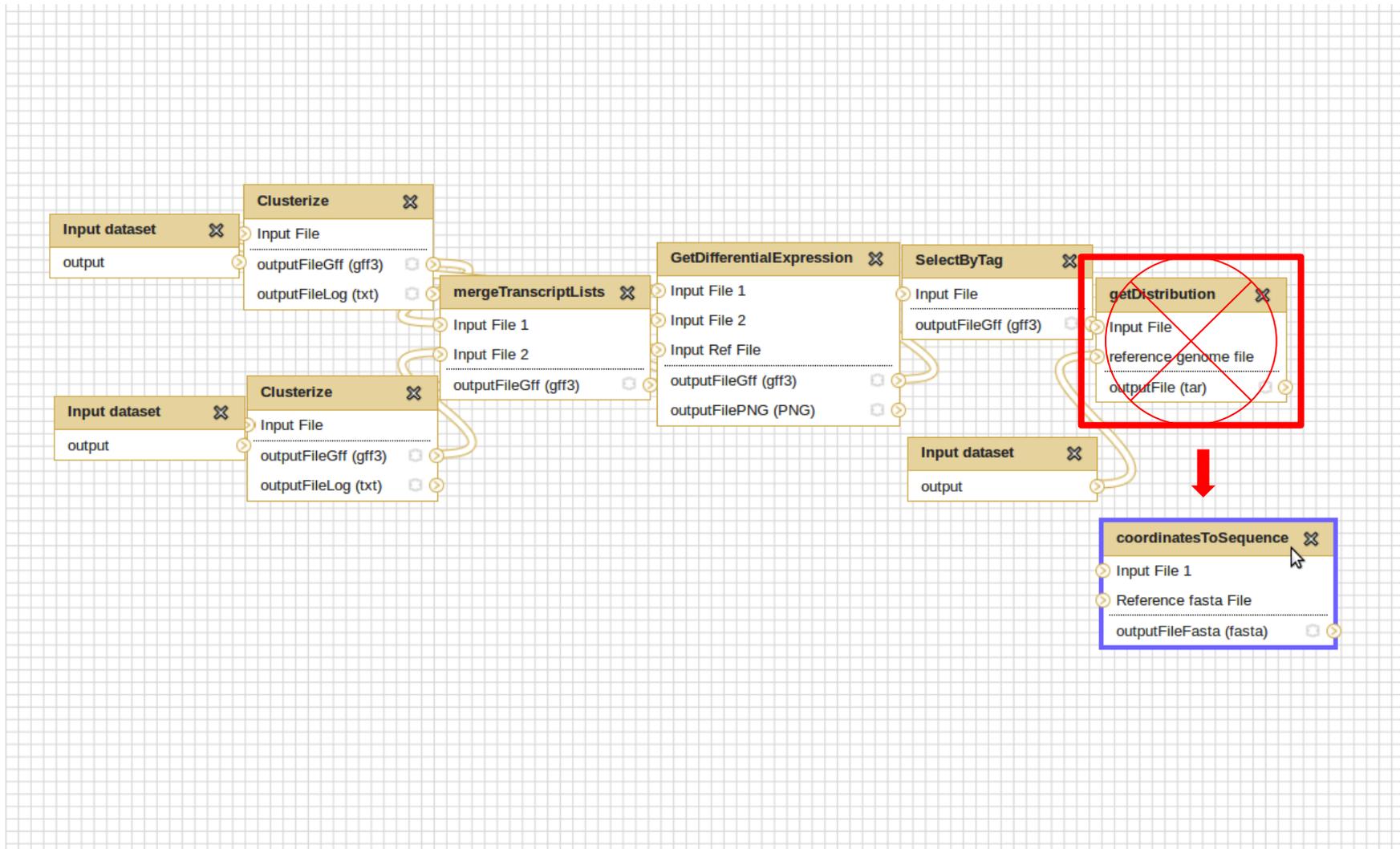
No workflows have

Other option

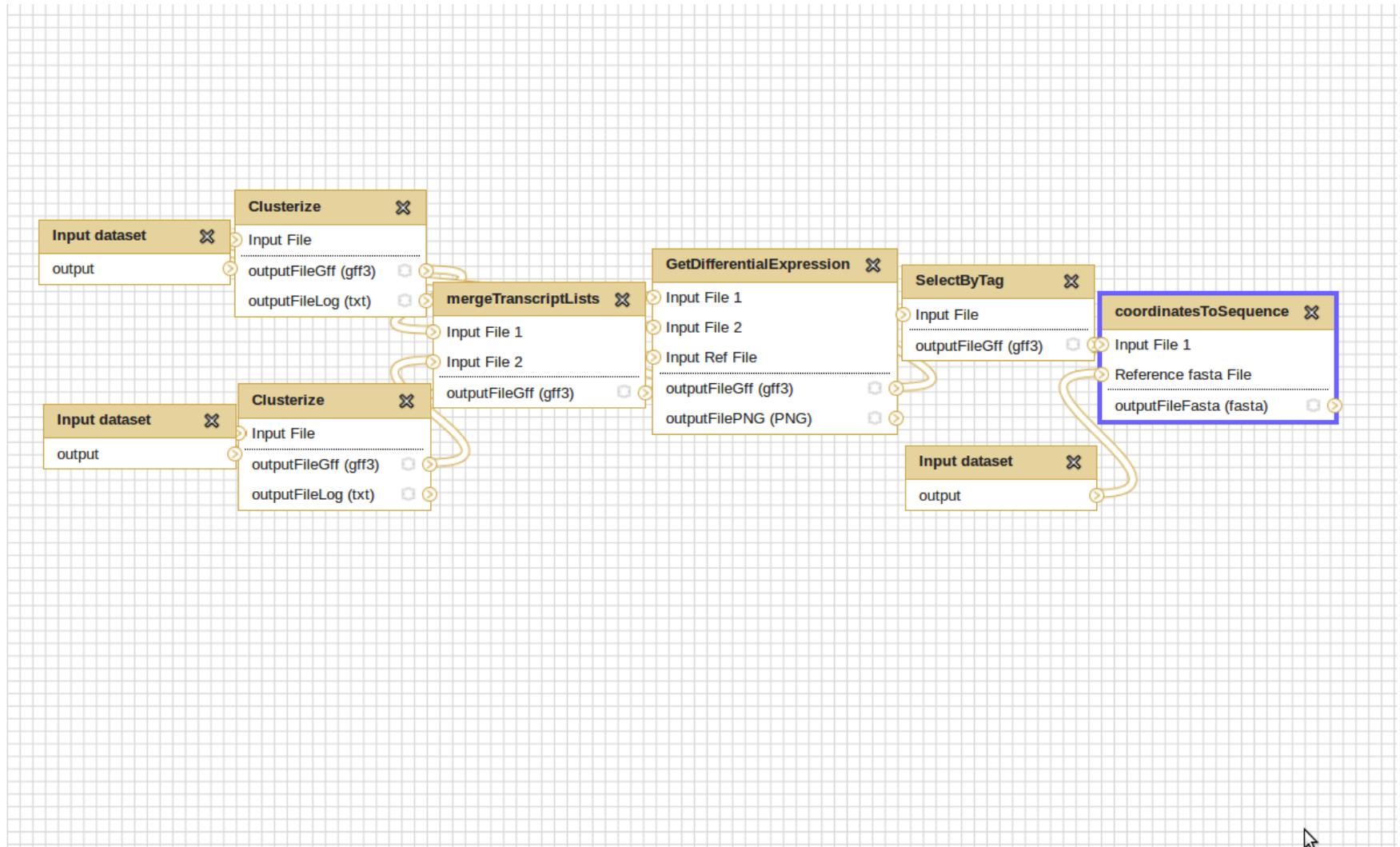
Configure your w



Edit S-MART workflow in Galaxy



Edit S-MART workflow in Galaxy





Export S-MART workflow from Galaxy

Galaxy Analyze Data **Workflow** Shared Data Admin Help User Using 5.9 Gb

Your workflows Create new workflow Upload or import workflow

Name		# of Steps
Differential	Edit	9
pipeLine_A	Run (uploaded file) ▾	14
pipeLine_C	Share or Publish (file) ▾	12
TRANS_ne	Download or Export (file) ▾	18
	Clone	
	Rename	
	Delete	

Workflow **shared with you by others**

No workflows have been shared with you.

Other options

Configure your workflow menu

Export S-MART workflow from Galaxy

Download or Export Workflow 'Differential_expression'

Download to File

[Download workflow to file so that it can be saved or imported into another Galaxy server.](#)

URL for Importing to Another Galaxy

[This workflow must be accessible before it can be imported into another Galaxy.](#)

Export to myExperiment

Export

myExperiment username:

myExperiment password:

[Back to Workflows List](#)



Import S-MART workflow in Galaxy

Galaxy Analyze Data **Workflow** Shared Data Admin Help User Using 2.1 Gb

Your workflows Create new workflow Upload or import workflow

Name	# of Steps
pipeLine_CIS ▾	13
TRANS_new ▾	18
pipeLine_ANTISENS ▾	14

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

Other options
Configure your workflow menu

Import S-MART workflow in Galaxy

Galaxy Analyze Data Workflow Shared Data Visualization Admin Help User Using 5.7 Gb

Import an exported Galaxy workflow file

URL for exported Galaxy workflow:

If the workflow is accessible via an URL, enter the URL above and click the **Import** button.

Exported Galaxy workflow file:

If the workflow is stored locally in a file, browse and select it and then click the **Import** button.

Envoi du fichier

yluo galaxy-local tools repet_pipe SMART DiffExpAnal

Raccourcis	Nom	Taille	Modifié
Rechercher	ecupse		22/02/2012
Récemment util...	Documents		02/12/2011
yluo	doc_ForGalaxy		10/11/2011
Bureau	Data_pair		10/11/2010
Système de fich...	Bureau		16:24
Documents	Biblio		20/12/2010
Musique	topcasedBookmarks.xml	1,3 Ko	17/01/2011
Images	runJobs.sh	675 octets	21/10/2011
Vidéos	runJob.py	373 octets	21/10/2011
Téléchargements	motivation_formation.odt	9,9 Ko	09/11/2011
Java	modifyImport.py	3,6 Ko	25/01/2011
	lSedCmdLines.sh	73,9 Ko	26/01/2011
	Galaxy-Workflow-TRANS_new.ga	23,6 Ko	15/02/2012
	Galaxy-Workflow-pipeLine_CIS.ga	17,5 Ko	15/02/2012
	Galaxy-Workflow-pipeLine_ANTISENS.ga	19,7 Ko	15/02/2012
	galaxy-dist-20111125.tar.gz	122,7 Mo	05/01/2012
	Galaxy8-[Annotation_File].gff3	182,9 Ko	10/01/2012

Tous les fichiers



Display S-MART workflow in Galaxy

Galaxy Analyze Data **Workflow** Shared Data Visualization Admin Help User Using 5.7 Gb

Your workflows

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

Name	# of Steps
pipeLine_ANTISENS (imported from uploaded file) ▼	13
pipeLine_CIS (imported from uploaded file) ▼	12
TRANS_new (imported from uploaded file) ▼	18

Workflows shared with you by others

No workflows have been shared with you.

Other options

[Configure your workflow menu](#)

Display S-MART workflow in Galaxy

Name	Owner	# of Steps	Show in menu
Differential_expression	You	9	<input checked="" type="checkbox"/>
pipeLine_ANTISENS (imported from uploaded file)	You	14	<input checked="" type="checkbox"/>
pipeLine_CIS (imported from uploaded file)	You	12	<input checked="" type="checkbox"/>
TRANS_new (imported from uploaded file)	You	18	<input checked="" type="checkbox"/>

Envoyer



Workflows

- [Differential_expression](#)
- [pipeLine_ANTISENS \(imported from uploaded file\)](#)
- [pipeLine_CIS \(imported from uploaded file\)](#)
- [TRANS_new \(imported from uploaded file\)](#)
- [All workflows](#)

Share S-MART workflow in Galaxy

Galaxy Analyze Data **Workflow** Shared Data Admin Help User Using 5.9 Gb

Your workflows Create new workflow Upload or import workflow

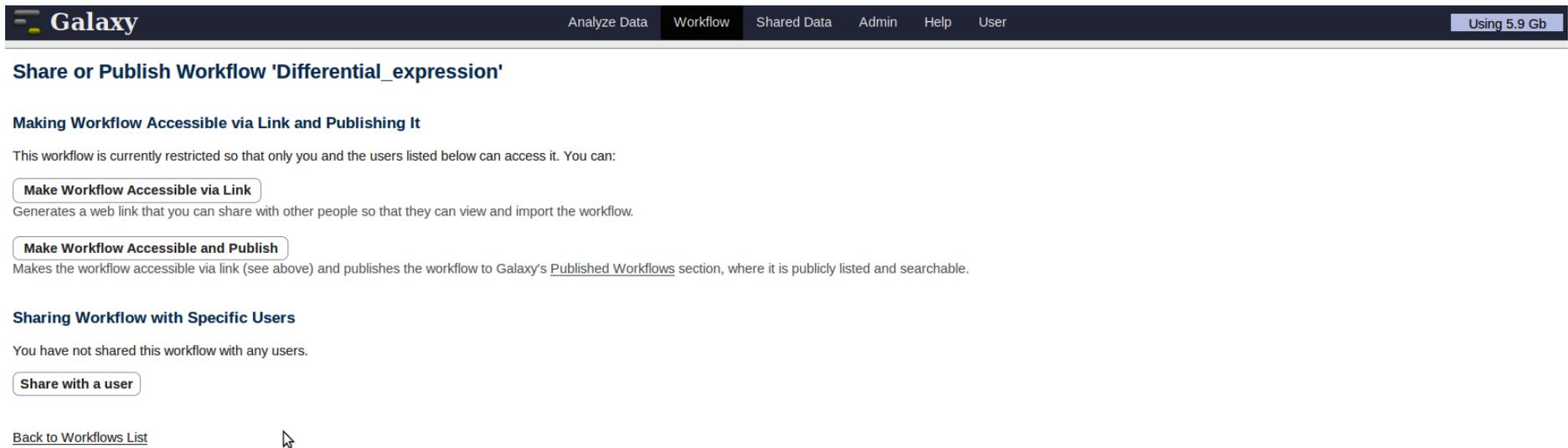
Name	# of Steps
Differential	9
pipeline_...	14
pipeline_...	12
TRANS_in...	18

Workflow actions: Edit, Run, **Share or Publish**, Download or Export, Clone, Rename, Delete

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

Other options
Configure your workflow menu

Share S-MART workflows in Galaxy



Galaxy Analyze Data Workflow Shared Data Admin Help User Using 5.9 Gb

Share or Publish Workflow 'Differential_expression'

Making Workflow Accessible via Link and Publishing It

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

- Make Workflow Accessible via Link**
Generates a web link that you can share with other people so that they can view and import the workflow.
- Make Workflow Accessible and Publish**
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.

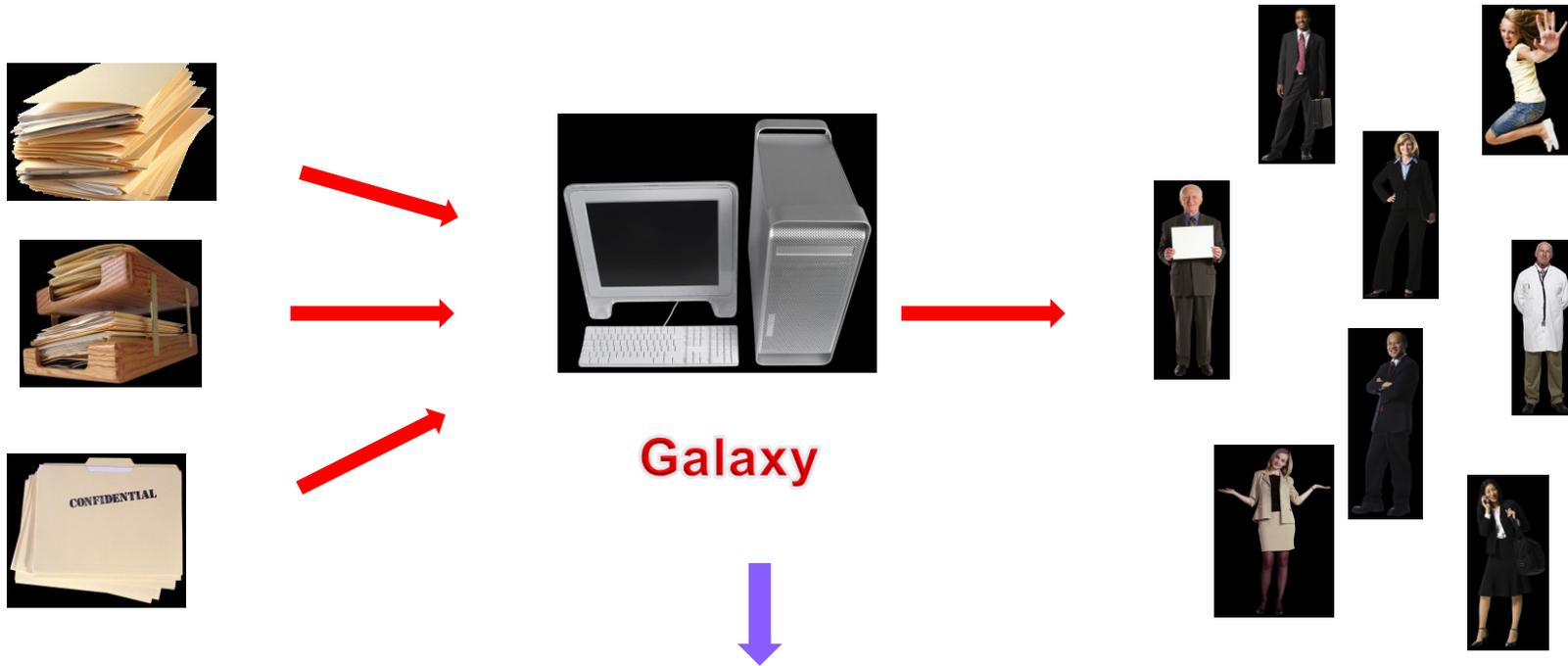
Sharing Workflow with Specific Users

You have not shared this workflow with any users.

- Share with a user**

[Back to Workflows List](#)

Share workflows & analysis tools in Galaxy



**APLIBIO Projects
(IBISA funding)**



APLIBIO

Posted on février 2, 2010 by [Ivan Moszer](#)

The APLIBIO regional center currently brings together 8 PFs located in Paris and in the Paris area with a staff of 80 people:

- [Curie](#)
- [eBio](#)
- [GENATLAS](#)
- [MicroScope](#)
- [MIGALE](#)
- [Pasteur](#)
- [RPBS](#)
- [URGI](#)

Biological models and applications. APLIBIO focuses its activities mainly on the following types of organism: microorganisms (*e.g.* infectious diseases at eBio and Pasteur, or environmental metagenomics at MicroScope and MIGALE), plants and their bioagressors such as fungi (URGI), and human (genetic diseases at GENATLAS, cancers at Curie). A specific activity is dedicated to protein structural analysis and chemoinformatics (RPBS and MIGALE).

Databases and software tools. All PFs are providing biological data collections, often combining heterogeneous data (genome sequences, expression data, biodiversity, genetic information, protein structures, etc.). Internationally recognized examples of such collections include the human gene and disease database GENATLAS, the plant genomics information system GnpIS or the microbial database GenoList. Elaborated software platforms are also developed, such as state-of-the-art annotation environments like MaGe and AGMIAL, or integrated web portals like MobyLe. These cutting-edge services are driven by specialized information technology developments (*e.g.* data integration, automated pipelines, text mining, database interoperability, human-computer interaction, etc.).

PARTNERS



EVENTS



TAGS

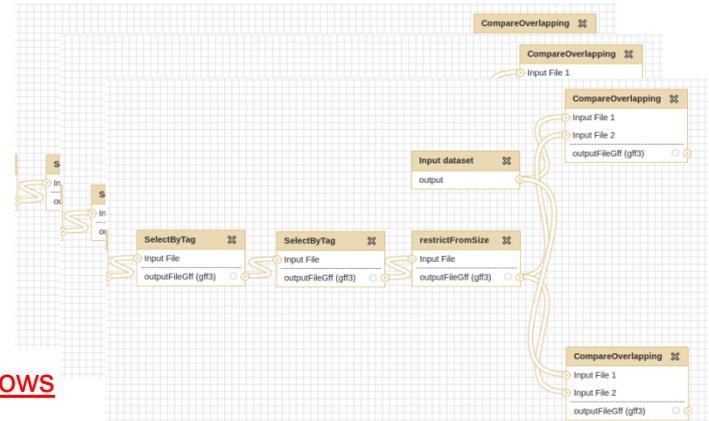
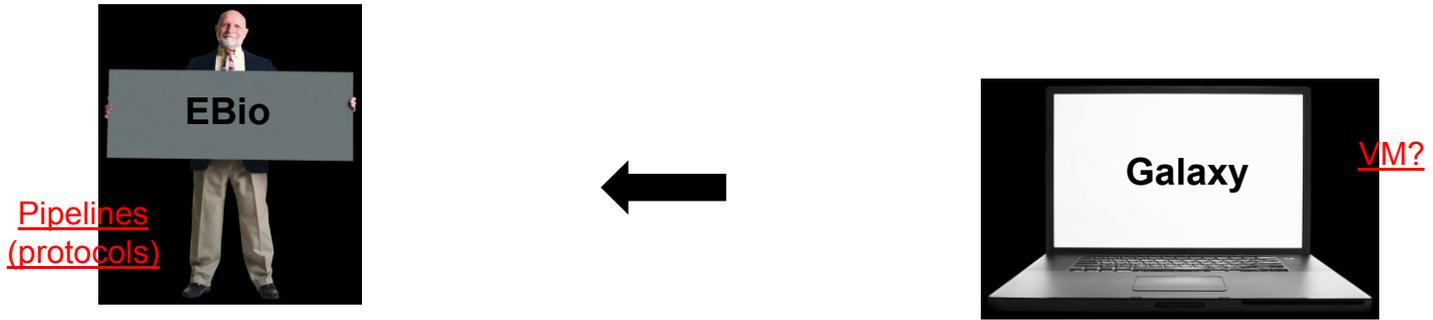
KEYWORDS

Computing **Databases** Data
distribution Genomics grid computing Intensive
computation Metagenomics online tool

APLIBIO Projects

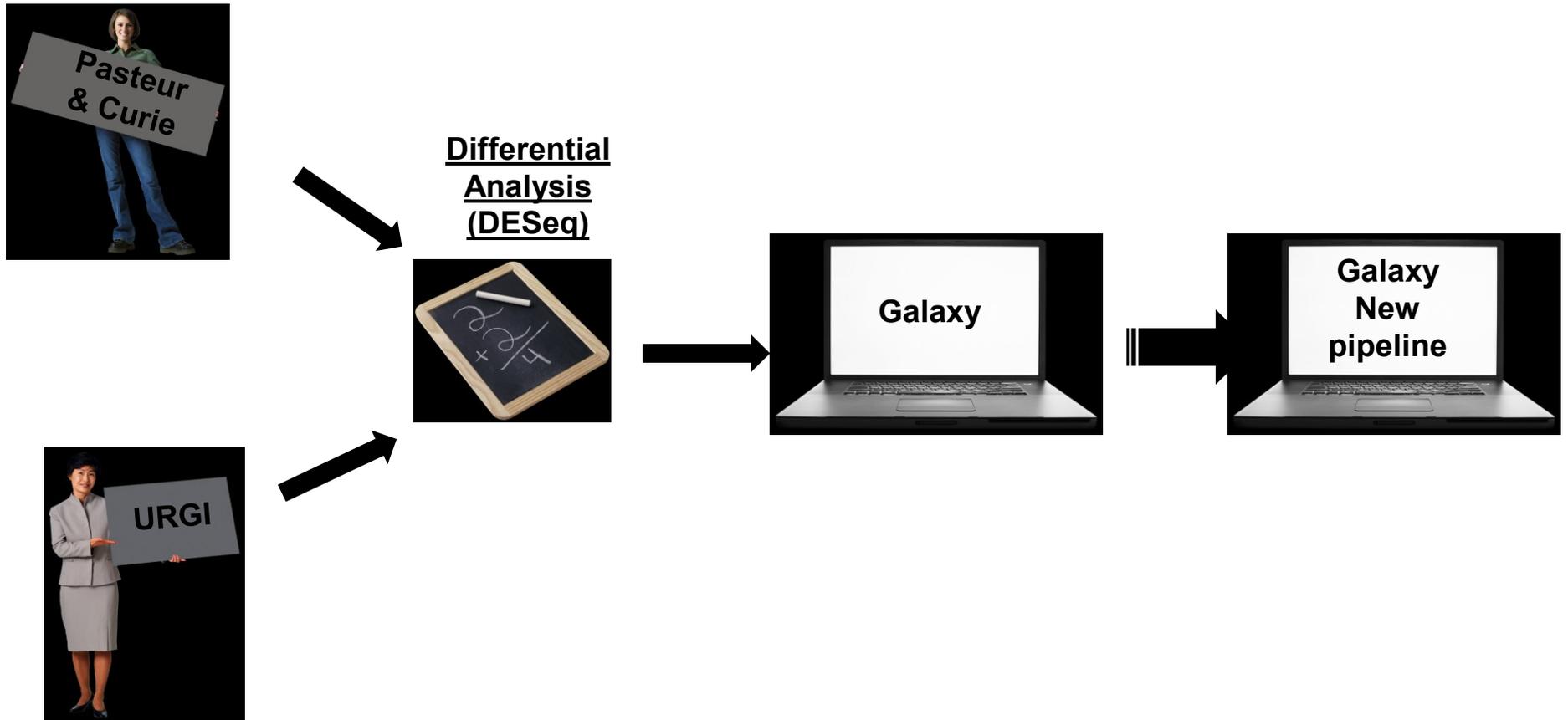


APLIBIO Projects



workflows

APLIBIO Projects



Acknowledgment



Inserm

Institut national
de la santé et de la recherche médicale

CECI



Inria

IBISA
INFRASTRUCTURES
BIOLÓGIE, SANTÉ
ET AGRICULTURE

Institut Pasteur

eBIO



MIG
Mathématique
Informatique
& Génome



Acknowledgment





Thank you for your attention !!!



Galaxy

Tool Intergration

- Easy to develop and add new tools but requires at least scripting skills
- Each tool is define into a xml file with a proper syntax.
[http://wiki.g2.bx.psu.edu/Admin/Tools/Tool Config Syntax](http://wiki.g2.bx.psu.edu/Admin/Tools/Tool%20Config%20Syntax)
- Once the tool is created, you need to declare it in Galaxy by updating.
`$GALAXY_DIR/galaxy-dist/tool_conf.xml`
- If you're using a job scheduler, you also need to declare the resources needed by your tool.
Update `$GALAXY_DIR/galaxy-dist/universe_wsgi.ini`
- Restart Galaxy
- If you plan to make change in any tool xml file, you will have to reload the tool configuration from the admin tab (no need to restart Galaxy)

Example of tool creation: XML file

- **Tool and command definition:**

```
<tool id="upload_file" name="Upload file">  
  <description>to current History</description>  
  <command interpreter="bash">upload_file.sh -b $file -o $out1</command>
```

- **Inputs definition:**

```
<inputs>  
  <param name="name" type="text" label="File Name"/>  
  <param name="extension" type="select" label="File type">  
    <option value="bam">Bam</option>  
    <option value="txt">Text</option>  
    <option value="fastq">Fastq</option>  
    <option value="csfasta">Csfasta</option>  
    <option value="qual">Qual</option>  
    <option value="bed">Bed</option>  
    <option value="gff">Gff</option>  
    <option value="pdf">Pdf</option>  
    <option value="vcf">VCF</option>  
    <option value="sam">Sam</option>  
    <option value="fasta">Fasta</option>  
    <option value="xsq">Xsq</option>  
  </param>  
  <param name="file" type="text" size="30" label="Path to file"></param>  
</inputs>
```

Example of tool creation: XML file

- Outputs definition:**

```

<outputs>
  <data format="bam" name="out1" label="{name}.{extension}">
    <change_format>
      <when input="extension" value="txt" format="txt" />
      <when input="extension" value="fastq" format="fastq" />
      <when input="extension" value="csfasta" format="csfasta" />
      <when input="extension" value="qual" format="qual" />
      <when input="extension" value="bed" format="bed" />
      <when input="extension" value="gff" format="gff" />
      <when input="extension" value="vcf" format="vcf" />
      <when input="extension" value="sam" format="sam" />
      <when input="extension" value="fasta" format="fasta" />
      <when input="extension" value="pdf" format="pdf" />
      <when input="extension" value="xsq" format="xsq" />
    </change_format>
  </data>
</outputs>
<help>
</help>
</tool>

```

Example of tool creation:

Execution file

(perl, python, shell, R, etc)

```
#!/bin/bash
```

```
while getopts "b:o:" optionName; do  
case "$optionName" in
```

```
  b) FILEORI="$OPTARG";;  
  o) FILEOUT="$OPTARG";;  
esac  
done
```

```
rm $FILEOUT
```

```
In -s $FILEORI $FILEOUT
```

Example of tool creation: Tool declaration

Edit `$GALAXY_DIR/galaxy-dist/tool_conf.xml`:

```
<?xml version="1.0"?>
<toolbox>
  <section name="Get Data" id="gettext">
    <tool file="data_source/upload.xml"/>
    <tool file="$PATH2NewTool/NewTool.xml" />
    <tool file="data_source/ucsc_tablebrowser.xml" />
    <tool file="data_source/ucsc_tablebrowser_test.xml" />
    <tool file="data_source/ucsc_tablebrowser_archaea.xml" />
    <tool file="data_source/bx_browser.xml" />
    <tool file="data_source/microbial_import.xml" />
    <tool file="data_source/biomart.xml" />
    <tool file="data_source/biomart_test.xml" />
  </section>
  ....
</toolbox>
```

Example of tool creation: Tool resources

Edit `$GALAXY_DIR/galaxy-dist/universe_wsgi.ini`:

```
# -- Database
```

```
# may use a SQLAlchemy connection string to specify an external database  
# instead. This string takes many options which are explained in detail in the  
# config file documentation.
```

```
#database_connection = sqlite:///./database/universe.sqlite?isolation_level=IMMEDIATE
```

```
database_connection = postgres://userName:password@urgiDB.versailles.inra.fr:port/userDatabase
```

Then, you need to restart your Galaxy instance to start working with your tool