

# Initiation à Galaxy, ressource pour le calcul en bioinformatique

## Application à la recherche de microsatellites et design de primers

Alexis Dereeper



Présentation extraite de session d'initiation à Galaxy South Green  
Alexis Dereeper, Jean-François Dufayard, Cédric Farcy, Ayité KOUGBEADJO, Pierre Larmande, Christine Tranchant  
avec l'aide des formateurs précédents: Félix Homa, Stéphanie Pointet, Marilyn Summo

# Galaxy

```
simmons@earth:~/work/xterm/orig/xterm-XF86-4.3.0$ ls
256colres.h      error.h          misc.c           Tekparse.h      VTPrsTbl.c
256colres.pl     fontutils.c     mkdirs.sh        Tekproc.c        vttests
88colres.h       fontutils.h     os2main.c        TekPrsTbl.c      wwidth.c
88colres.pl      Imakefile       plink.sh         tektests         wwidth.h
AAA_README_VMS.txt input.c         precompose.c     termcap          xcharmouse.h
aclocal.m4       INSTALL         precompose.h     terminfo         xstrings.c
button.c         charclass.c     keysym2ucs.c     testxmc.c       xstrings.h
charclass.h      charclass.h     link_xp.com      trace.c          xterm_arp.opt
charproc.c       charsets.c      main.c           trace.h          xtermcfg.hin
charsets.c       config.guess    main.h           unicode          XTerm-col.ad
config.sub       make.com        README           util.c           xterm.dat
configure        Makefile        README.os390     uxterm           xterm.h
configure.in     Makefile.in     screen.c         version.h        xterm_io.h
cursor.c         data.c          scrollbar.c       vms.c           xterm.log.html
data.c           data.h          sinstall.sh      vms.h           xterm.man
doublechr.c      menu.h          tabs.c           VTparse.def      xutf8.c
doublechr.c      mininstall.sh   Tekparse.def     VTparse.h        xutf8.h
```

Communication



Communication



# Galaxy

The screenshot displays the Galaxy web interface at [galaxy.southgreen.fr/galaxy/](http://galaxy.southgreen.fr/galaxy/). The browser's address bar shows the URL, and the search bar contains "southgreen". The interface features a dark navigation bar with the Galaxy logo and menu items: Analyze Data, Workflow, Shared Data, Visualization, Admin, Help, and User. A status bar on the right indicates "Using 51%".

**Tools Panel (Left):** A sidebar with a search bar and a list of tool categories and specific tools, including:

- Get Data
- Send Data
- BASIC TOOLS
  - Text Manipulation
  - Filter and Sort
  - Join, Subtract and Group
  - Convert Formats
- SEQUENCE ANALYSIS
  - EMBOSS
  - Operate on Genomic Intervals
  - Fetch Sequences
  - Genomics
  - Fetch Alignments
  - Extract Features
- NGS ANALYSIS
  - NGS: Quality Control
  - NGS : Mapping
  - NGS: GATK Tools
  - NGS: GATK2 Tools
  - NGS: SAM/BAM Manipulations
  - NGS: RNASeq
  - NGS: Assembly
  - NGS: Small RNAs
  - Bedtools
  - Picard Tools
- SNP ANALYSIS
  - NGS: SNP Calling
  - Varscan
  - Population structure
  - GWAS
  - VCFtools

**Main Content Area:** The central panel displays the "South Green bioinformatics platform" logo and a welcome message: "Welcome to GALAXY ... at your disposal as part of the services provided by SouthGreen". A yellow box contains a message: "In order to figure out which tools were made available by our team, please activate the 'tool search' functionality from the Options drop-down and type 'southgreen' in the lookup filter." Below this, a blue box states: "Requests for making new tools available shall be addressed to [galaxy-dev-southgreen@cirad.fr](mailto:galaxy-dev-southgreen@cirad.fr)". At the bottom, a small text line reads: "The GALAXY project is supported in part by INSE, NHGRI, and the Huck Institutes of the Life Sciences."

**History Panel (Right):** A sidebar showing a list of datasets and workflows. The top section is titled "History" and includes a search bar. The list includes:

- SNIPlay (97 shown, 618 deleted, 40 hidden, 11.72 GB)
- 751: vcf output
- 747: vcf output
- 746: vcf output
- 733: analyse.log
- 732: analyse.mds\_plot.txt
- 731: analyse.ibs\_matrix.txt
- 725: All Logs
- 724: All Outputs
- 723: Best K Logfile
- 722: Best K Groups
- 721: Best K Output
- 716: densities.by\_sample
- 715: densities
- 713: snp\_density.windowed.pi
- 712: snp\_density.TsTv

# Galaxy

Accès direct au cluster

Interface et compte utilisateur indépendants du poste client

Centralisation et partage des démarches

Centralisation et partage des données

Utilisable depuis n'importe quel poste connecté à internet

The screenshot displays the Galaxy web interface. The top navigation bar includes links for 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. The main content area features the 'SouthGreen bioinformatics platform' logo and a welcome message. A central box provides instructions on how to activate the 'tool search' functionality. Below this, a box mentions that requests for making new tools should be addressed to [galaxy-dev-southgreen@cirad.fr](mailto:galaxy-dev-southgreen@cirad.fr). The right sidebar shows a 'History' section with a search bar and a list of datasets, including 'SNIPlay', '751: vcf output', '747: vcf output', '746: vcf output', '733: analyse.log', '732: analyse.mds\_plot.txt', '731: analyse.ibs\_matrix.txt', '725: All Logs', '724: All Outputs', '723: Best K Logfile', '722: Best K Groups', '721: Best K Output', '716: densities.by\_sample', '715: densities', '713: snp\_density.windowed.pi', and '712: snp\_density.TsTv'.



# Galaxy - Vue générale

The screenshot shows the Galaxy web interface at [galaxy.southgreen.fr/galaxy/](http://galaxy.southgreen.fr/galaxy/). The browser address bar shows the URL and a search for "southgreen". The interface includes a top navigation bar with tabs: Analyze Data, Workflow, Shared Data, Visualization, Admin, Help, and User. A status bar on the right indicates "Using 51%".

The left sidebar, titled "Tools", is circled in red and contains a search bar and a list of tool categories:

- Get Data
- Send Data
- BASIC TOOLS
  - Text Manipulation
  - Filter and Sort
  - Join, Subtract and Group
  - Convert Formats
- SEQUENCE ANALYSIS
  - EMBOSS
  - Operate on Genomic Intervals
  - Fetch Sequences
  - Genomics
  - Fetch Alignments
  - Extract Features
- NGS ANALYSIS
  - NGS: Quality Control
  - NGS : Mapping
  - NGS: GATK Tools
  - NGS: GATK2 Tools
  - NGS: SAM/BAM Manipulations
  - NGS: RNASeq
  - NGS: Assembly
  - NGS: Small RNAs
  - Bedtools
  - Picard Tools
- Population structure
- GWAS
- VCFtools

The main content area displays the "South Green bioinformatics platform" logo and a welcome message: "Welcome to GALAXY ... at your disposal as part of the services provided by SouthGreen". It includes a message about activating the "tool search" functionality and a contact email for tool requests: [galaxy-dev-southgreen@cirad.fr](mailto:galaxy-dev-southgreen@cirad.fr).

The right sidebar, titled "History", is also circled in red and shows a list of datasets:

- 751: vcf output
- 747: vcf output
- 746: vcf output
- 733: analyse.log
- 732: analyse.mds\_plot.txt
- 731: analyse.ibs\_matrix.txt
- 725: All Logs
- 724: All Outputs
- 723: Best K Logfile
- 721: Best K Output
- 716: densities.by\_sample
- 715: densities
- 713: snp\_density.windowed.pi
- 712: snp\_density.TsTv

At the bottom right, a red box contains the text "Données".

At the bottom left, a red box contains the text "Boîte à outils".

# Galaxy - Vue générale

Chaque programme peut être configuré précisément

**Galaxy** Analyze Data Workflow Shared Data Visualization Admin Help User Using 51%

**Tools**

- [Convert Formats](#)
- SEQUENCE ANALYSIS**
- [EMBOSS](#)
- [Operate on Genomic Intervals](#)
- [Fetch Sequences](#)
- [Genomics](#)
- [Fetch Alignments](#)
- [Extract Features](#)
- NGS ANALYSIS**
- [NGS: Quality Control](#)
- [NGS : Mapping](#)
- [Bowtie2](#) - Map reads against reference genome
- [Map with BWA for Illumina](#)
- [Map with BWA for SOLiD](#)
- [TopHat](#) Gapped-read mapper for RNA-seq data
- [Megablast](#) compare short reads against htgs, nt, and wgs databases
- [Map with BWA-MEM](#) - map medium and long reads (> 100 bp) against reference genome
- [Map with BWA](#) - map short reads (< 100 bp) against reference genome
- [NGS: GATK Tools](#)
- [NGS: GATK2 Tools](#)
- [NGS: SAM/BAM Manipulations](#)
- [NGS: RNASeq](#)
- [NGS: Assembly](#)
- [NGS: Small RNAs](#)
- [Bedtools](#)

**Map with BWA** - map short reads (< 100 bp) against reference genome (Galaxy Version 0.3.1) Options

Will you select a dataset to use as a reference genome?

Use a built-in genome index

Built-in genomes are indexed using default options. See `Indexes` section of help below

**Using reference genome**

Musa acuminata pseudo-chromosome (version 1.0)

Select genome from the list

**Select input type**

Paired fastq

Select between fastq and bam datasets and between paired and single end data

**Select first set of reads**

No fastqsanger dataset available.

Specify dataset with forward reads

**Select second set of reads**

No fastqsanger dataset available.

Specify dataset with reverse reads

**Set advanced paired end options?**

Do not set

Provides additional controls

**Set read groups information?**

Do not set

Specifying read group information can greatly simplify your downstream analyses by allowing combining multiple datasets.

**Select analysis mode**

1.Simple Illumina mode

Execute

**What is does**

BWA is a software package for mapping low-divergent sequences against a large reference genome, such as the human genome. The bwa-aln algorithm is designed for Illumina sequence reads up to 100bp. For longer reads use BWA-MEM algorithm distributed as a separate Galaxy tool.

This Galaxy tool wraps bwa-aln, bwa-samse and -sampe modules of bwa read mapping tool:

**History**

search datasets

**Uniqprimer**  
18 shown, 81 deleted  
20.48 MB

**99: Logfile**

**98: Fasta for Primer3**

**97: Primer list**

**93: Logfile**

**92: Fasta for Primer3**

**91: Primer list**

**90: Logfile**

**89: Fasta for Primer3**

**88: Primer list**

**41: Logfile**

**40: Fasta for Primer3**

**39: Primer list**

**28: Logfile**

**27: Output**

**4:**  
PX099A\_sequence.fa

**3:**  
Xoo\_KACC\_10331.fasta

**2:**  
Xoc\_BLS256.fasta

**Configuration d'un programme**

# Galaxy - accès aux données

Les données peuvent être chargées depuis votre ordinateur local

The screenshot displays the Galaxy web interface. The top navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Admin', 'Help', and 'User'. The left sidebar shows 'Tools' with a search bar and a list of 'Recently Used' tools. The 'Upload File (version 1.1.3)' tool is selected, showing its configuration options. Three red annotations highlight key features:

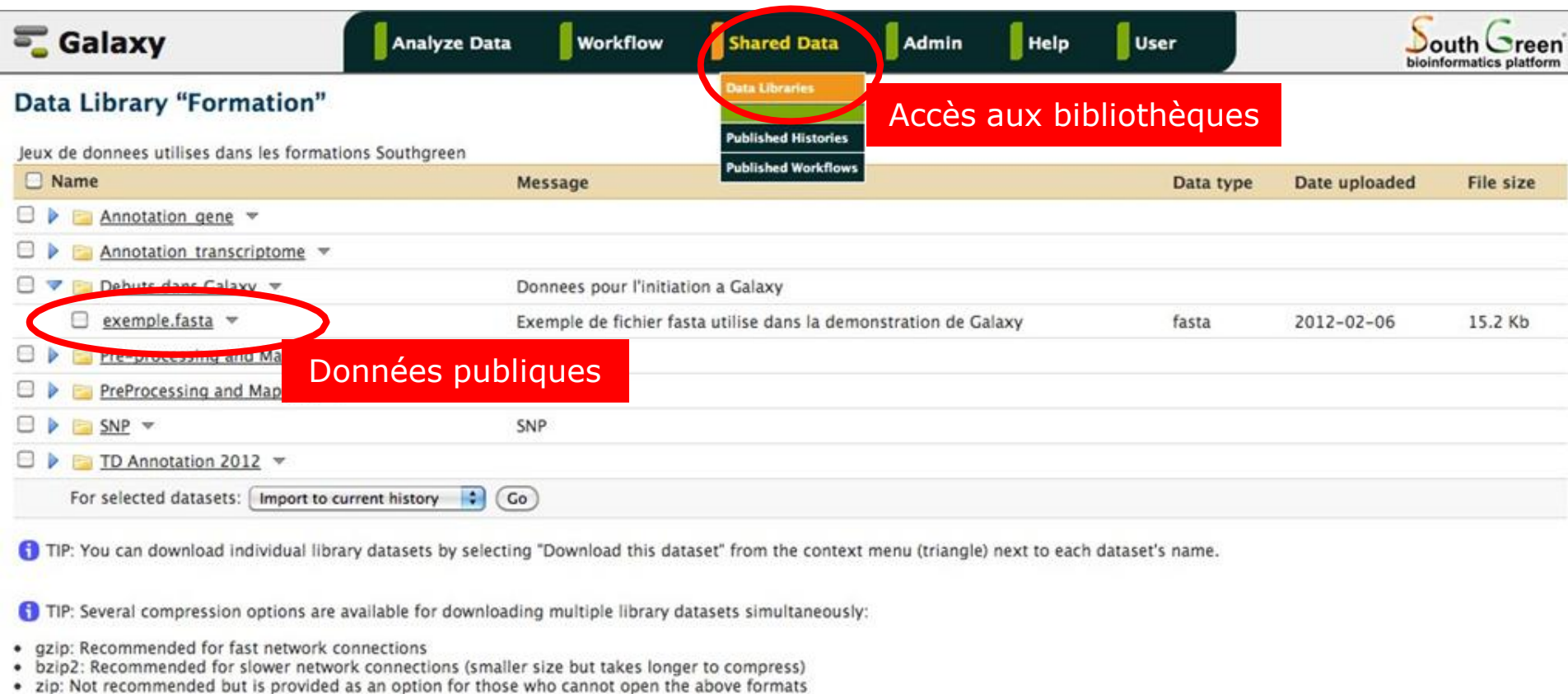
- Pour importer depuis votre ordinateur local:** Points to the 'Upload File from your computer' option in the 'Recently Used' list and the 'Choisissez un fichier' button in the 'File' section.
- Pour importer des données depuis une autre page web ou par copier/coller:** Points to the 'URL/Text' input field and the explanatory text below it.
- Explication sur les différents formats:** Points to the 'Auto-detect' section, which lists various file formats like Axt, Fasta, Fastq, etc.

The right sidebar shows a 'History' panel with a list of previous jobs, including '20: PhyML on data 18', '19: PhyML on data 18', and '14: new.fasta'.



# Galaxy - accès aux données

Les données peuvent être chargées depuis des bibliothèques partagées



The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data' (highlighted with a red circle), 'Admin', 'Help', and 'User'. A dropdown menu for 'Shared Data' is open, showing 'Data Libraries', 'Published Histories', and 'Published Workflows'. A red box labeled 'Accès aux bibliothèques' points to the 'Data Libraries' option. Below the navigation bar, the 'Data Library "Formation"' section is visible. It contains a table of datasets. The table has columns: 'Name', 'Message', 'Data type', 'Date uploaded', and 'File size'. The dataset 'exemple.fasta' is highlighted with a red circle, and a red box labeled 'Données publiques' points to it. At the bottom, there are tips and a list of compression options.

Galaxy

Analyze Data Workflow **Shared Data** Admin Help User

South Green  
bioinformatics platform

Data Library "Formation"

Jeux de données utilisés dans les formations Southgreen

<input type="checkbox"/> Name	Message	Data type	Date uploaded	File size
<input type="checkbox"/> <a href="#">Annotation_gene</a> ▾				
<input type="checkbox"/> <a href="#">Annotation_transcriptome</a> ▾				
<input type="checkbox"/> <a href="#">Debut dans Galaxy</a> ▾	Données pour l'initiation à Galaxy			
<input type="checkbox"/> <a href="#">exemple.fasta</a> ▾	Exemple de fichier fasta utilisé dans la démonstration de Galaxy	fasta	2012-02-06	15.2 Kb
<input type="checkbox"/> <a href="#">Pre-processing and Mapping</a> ▾				
<input type="checkbox"/> <a href="#">PreProcessing and Mapping</a> ▾				
<input type="checkbox"/> <a href="#">SNP</a> ▾	SNP			
<input type="checkbox"/> <a href="#">TD Annotation 2012</a> ▾				

For selected datasets:

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

**TIP:** Several compression options are available for downloading multiple library datasets simultaneously:

- gzip: Recommended for fast network connections
- bzip2: Recommended for slower network connections (smaller size but takes longer to compress)
- zip: Not recommended but is provided as an option for those who cannot open the above formats



# Galaxy - Historiques des analyses

La notion d'historiques multiples est très importante pour gérer ses résultats

The screenshot shows the 'History' sidebar on the left with a list of analysis histories. A context menu is open over the sidebar, listing various actions for managing histories.

**History**

Options ▾

Click to rename history

Unnamed history 4.0 Mb

- 20: PhyML on data 18
- 19: PhyML on data 18
- 18: Fasta2Phylip on data 16
- 17: Gblocks on data 15
- 16: Gblocks on data 15
- 15: MAFFT on data 14
- 14: new.fasta

**History Lists**

- Saved Histories
- Histories Shared with Me
- Current History
- Create New
- Clone
- Copy Datasets
- Share or Publish
- Extract Workflow
- Dataset Security
- Show Deleted Datasets
- Show Hidden Datasets
- Purge Deleted Datasets
- Show Structure
- Export to File
- Delete
- Delete Permanently
- Other Actions
- Import from File

Historique courant

## Saved Histories

search history names and tags

Advanced Search

<input type="checkbox"/>	Name	Datasets	Tags	Sharing	Size on Disk	Created	Last Updated
<input type="checkbox"/>	Formation fevrier ▾	28	0 Tags		0 bytes	Feb 09, 2012	less than a minute ago
<input type="checkbox"/>	Analyse NSLTps ▾	11	0 Tags		98 bytes	Aug 22, 2011	1 minute ago
<input type="checkbox"/>	Unnamed history ▾	7	0 Tags		4.0 Mb	Apr 18, 2012	Apr 23, 2012
<input type="checkbox"/>	Clone of 'Unnamed history' shared by 'dominique.this@supagro.inra.fr' (active items only) ▾	150	1 0 Tags		56.1 Mb	Apr 23, 2012	Apr 23, 2012
<input type="checkbox"/>	Unnamed history ▾	3	0 Tags		0 bytes	Feb 09, 2012	Feb 09, 2012

Listes des historiques personnels, et publiés

The screenshot shows the 'Published Histories' section of the Galaxy platform. It includes a search bar, a table of published histories, and a sidebar with navigation links.

**Galaxy** Analyze Data Workflow Shared Data Admin

**Published Histories**

search name, annotation, owner, and tag

Advanced Search

Name	Annotation	Owner	Community Rating	Community
RC2		rogeriomercers	★★★★★	
RC2 Formation28		reynaldo	★★★★★	
imported: RC4 workflow		mruiZ	★★★★★	
RC3 alexis		mruiZ	★★★★★	
Unnamed history		formation12	★★★★★	

Data Libraries  
Published Histories  
Published Workflows

# Galaxy - Historiques des analyses

**South Green**  
bioinformatics platform

**History Lists**

- Saved Histories
- Histories Shared with Me
- Current History
- Create New
- Clone
- Copy Datasets
- Share or Publish
- Extract Workflow
- Dataset Security
- Show Deleted Datasets
- Show Hidden Datasets
- Purge Deleted Datasets
- Show Structure
- Export to File
- Delete
- Delete Permanently
- Other Actions
- Import from File

**Saved Histories**

search history names  
Advanced Search

**Cliquez ici pour afficher le panneau de gestion des historiques**

<input type="checkbox"/>	Name	Datasets	Tags	Sharing	Size on Disk	Created	Last Updated	Status
<input type="checkbox"/>			0 Tags		0 bytes	less than a minute ago	less than a minute ago	current history
<input type="checkbox"/>			0 Tags		0 bytes	3 days ago	~ 4 hours ago	
<input type="checkbox"/>	IGC3 combine	4	0 Tags		590.0 Mb	~ 6 hours ago	~ 5 hours ago	
<input type="checkbox"/>	Iq010 Clean Data	4	1 0 Tags		38.6 Gb	Jun 04, 2012	3 days ago	
<input type="checkbox"/>	Iq09 Clean Data	4	1 0 Tags		38.0 Gb	Jun 04, 2012	3 days ago	
<input type="checkbox"/>	Iq08 Clean Data	4	1 0 Tags		28.9 Gb	Jun 04, 2012	3 days ago	
<input type="checkbox"/>	Iq07 Clean Data	4	1 0 Tags		25.4 Gb	Jun 04, 2012	3 days ago	
<input type="checkbox"/>	Iq06 Clean Data	5	0 Tags		26.9 Gb	Jun 04, 2012	3 days ago	
<input type="checkbox"/>	Iq05 Clean Data	4	1 0 Tags		31.9 Gb	Jun 04, 2012	3 days ago	
<input type="checkbox"/>	Iq04 Clean Data	4	1 0 Tags		35.2 Gb	Jun 04, 2012	3 days ago	
<input type="checkbox"/>	Iq03 Clean Data	4	1 0 Tags		27.7 Gb	Jun 04, 2012	3 days ago	
<input type="checkbox"/>	Iq02 Clean Data	4	1 0 Tags		28.6 Gb	Jun 04, 2012	3 days ago	
<input type="checkbox"/>	Iq01 Clean Data	4	1 0 Tags		30.3 Gb	May 31, 2012	3 days ago	
<input type="checkbox"/>	arcad_test_data	1	1 0 Tags		2.6 Gb	Apr 24, 2012	Jun 11, 2012	
<input type="checkbox"/>	Cleaning NGS felix	1	1 0 Tags			May 07, 2012	May 07, 2012	
<input type="checkbox"/>	galaxy_test	1	1 0 Tags			May 04, 2012	May 07, 2012	

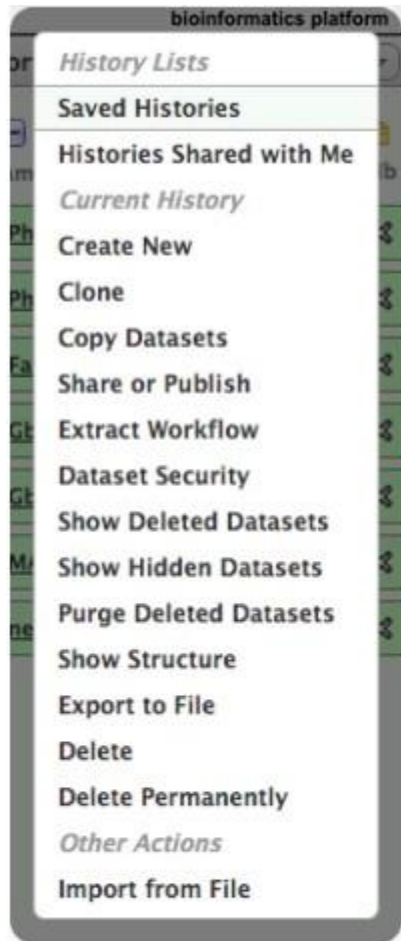
**Selectionnez les historiques à supprimer**

**Suppression temporaire**

**Suppression définitive**

For 0 selected histories:

# Galaxy - Historique des analyses



Visualisation du fichier

Modification des attributs du fichier

Suppression du fichier

La notion d'historiques multiples est très importante pour gérer ses résultats

Ne jamais utiliser de caractères accentués dans le nom des fichiers ou historiques ainsi que dans les annotations.

# Galaxy - Boîte à outils



Les outils sont (trop ?) nombreux,  
et classés par catégories

Outils testés et maintenus par Southgreen

- La version originale de Galaxy est livrée avec des dizaines d'outils
- La plateforme South Green a ajouté des dizaines d'outils supplémentaires
- Un outil de recherche par mot clef est disponible.

Outils installés sur le Galaxy original



# Galaxy - Workflows

Un workflow est un enchaînement d'outils paramétrés, ils sont partageables et publiables

The screenshot displays the Galaxy Workflow Canvas interface. The top navigation bar includes links for Analyze Data, Workflow, Shared Data, Admin, Help, and User. The left sidebar lists various tools categorized under 'Tools' and 'UNTESTED TOOLS'. The main canvas area shows a workflow titled 'Greenphyl3\_aminoacids' with a grid of tool nodes connected by lines representing data flow. The tools include 'Input dataset', 'Switch\_alignment', 'MAFFT', 'Fasta2Phylip', 'PhyML', 'HMMBUILD', and 'HMMCALIBRATE'. A red box labeled 'Canvas d'édition' points to the workflow grid. Another red box labeled 'Paramétrage' points to the configuration panel on the right, which shows settings for the 'PhyML' tool, such as 'Alignment in phylip format', 'Data type: Amino acids', 'Evolution model: LG', and 'Discrete gamma model: Use a gamma model'.

Galaxy

Analyze Data Workflow Shared Data Admin Help User

Tools Options Workflow Canvas | Greenphyl3\_aminoacids Options Det

search tools

Get Data  
Send Data

TOOLS

Convert Formats  
Evolution  
ESTik  
Filter and Sort  
Gene/Protein prediction  
SAT  
NGS: Quality Control  
NGS: Mapping  
NGS: SAM/BAM Manipulations  
NGS: SNP Detection  
Protein Structures  
Sequence comparisons

UNTESTED TOOLS

Text Manipulation  
Filter and Sort  
Join, Subtract and Group  
Convert Formats  
Extract Features  
Fetch Sequences  
Fetch Alignments  
Operate on Genomic Intervals  
Statistics  
Graph/Display Data  
Multiple regression  
Multivariate Analysis  
Evolution  
Metagenomic analyses

Canvas d'édition

Paramétrage

Tool: PhyML

Alignment in phylip format  
Data input 'input' (phylip)

Data type:  
Amino acids

Evolution model: ▼  
LG

Discrete gamma model:  
Use a gamma model

Number of categories for the discrete gamma model: ▼  
4

Shape parameter of the gamma model:  
e

Branch support:  
SH-like aLRT

Proportion of invariant sites: ▼  
0.0

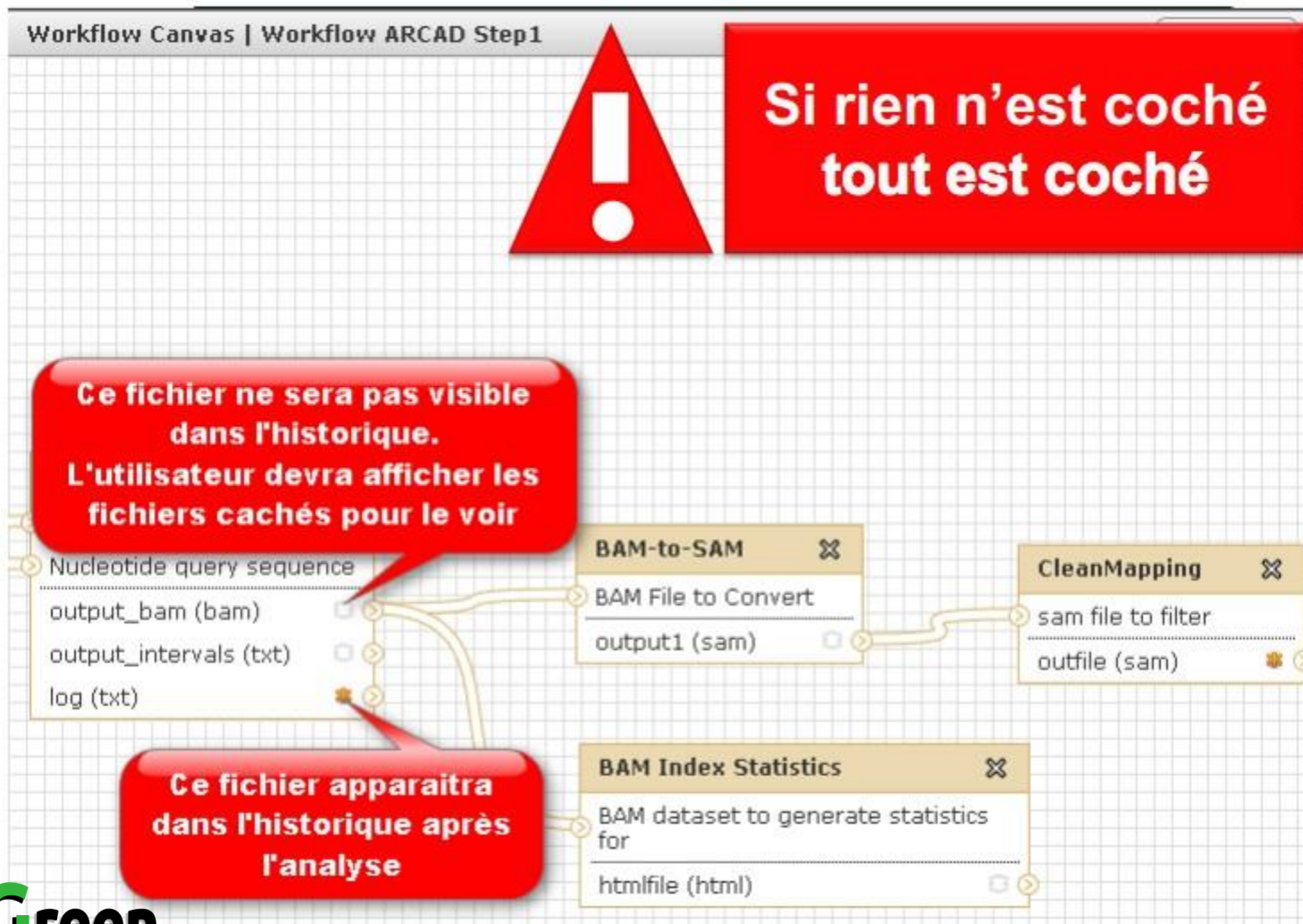
Tree topology search operation: ▼  
Best of NNI and SPR

Random starting points:  
Don't add random starting points

Edit Step Actions

# Galaxy - Workflows

Lors de la construction d'un workflow vous pouvez décider des fichiers qui seront visibles dans l'historique après l'analyse.



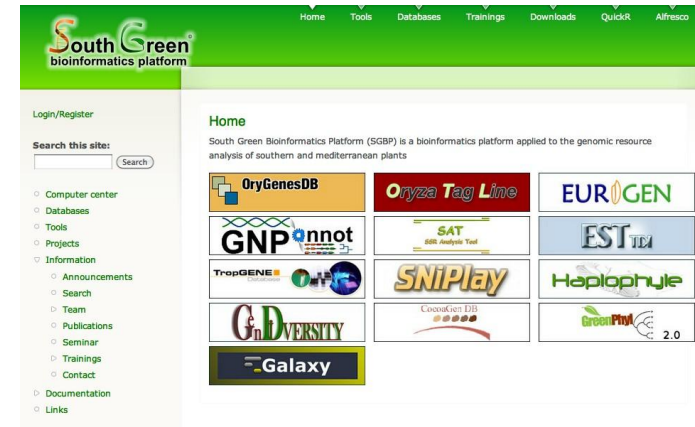
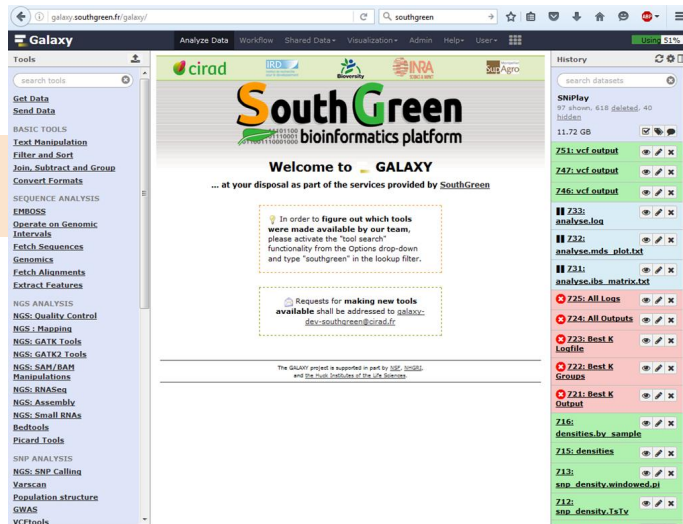
# Galaxy, sa place dans la plateforme

## Reproductibilité des analyses

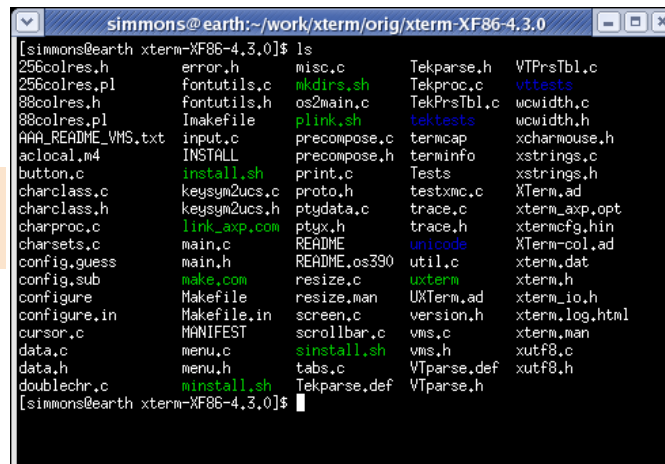
### Données à analyser

### Données analysées et explorables

### Utilisateur averti



### Utilisateur expert



## Séances de travail "développeur"

- 1 après-midi toutes les 2 semaines.
- Entre 8 et 12 bioinformaticiens, développeurs, biologistes.
- Travail en binômes pour faire évoluer la plateforme (Mise à jour de briques, documentation, tests et résolutions de problèmes, implémentation de nouvelles fonctionnalités...).

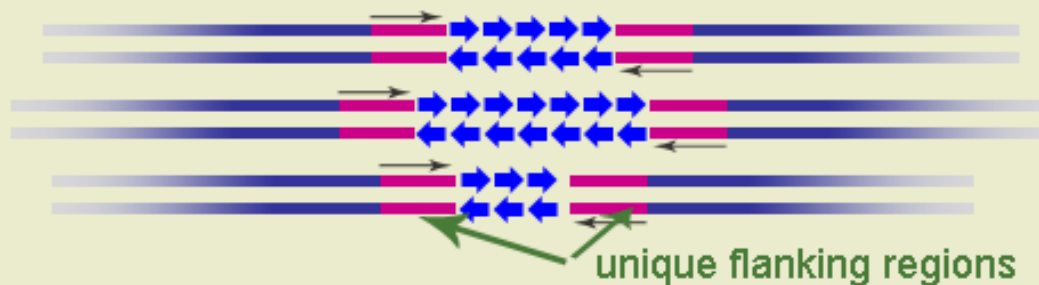


# TD: Application à la recherche de microsatellites

# Marqueurs Microsatellites

- Microsatellites = SSR (Simple Sequence Repeat)
- Variabilité du nombre de courtes répétitions en tandem. Répétitions de 1 à 13 pb (ex: (A)<sub>12</sub>, (AC)<sub>8</sub>)
- Haute variabilité
- Le nombre d'allèles n'est pas connu
- Marqueurs dispersés dans le génome  
~2% du génome, dinucléotide ~ 0.5%
- Utilisé comme marqueur génétique (marqueur de maladie)

The number of SSRs is highly variable among individuals



# Marqueurs Microsatellites

- ⇒ Utilisation de primers spécifiques des régions flanquantes pour l'amplification et l'analyse de génotypage
- ⇒ Les primers permettront de produire pour chaque individu de l'espèce des tailles de produits d'amplification différentes selon le nombre de répétition

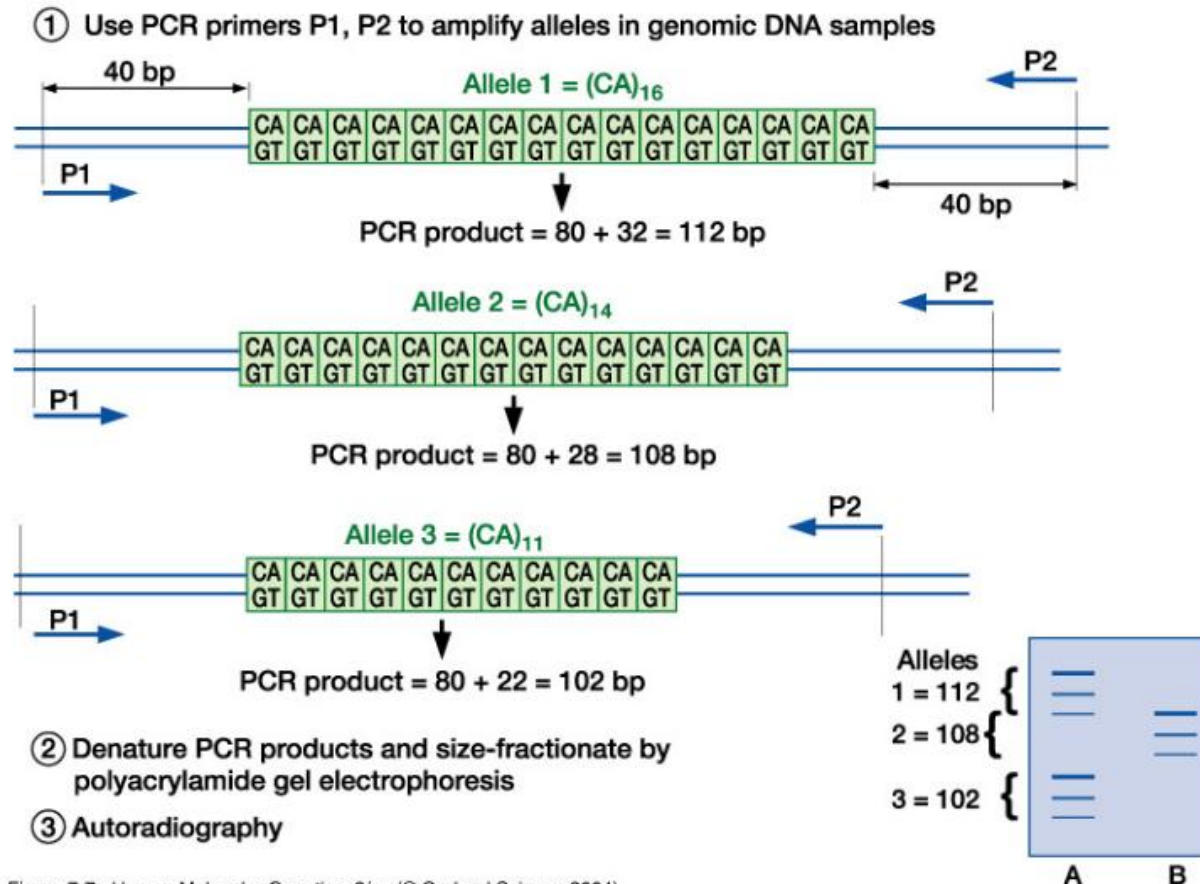


Figure 7-7 Human Molecular Genetics, 3/e. (© Garland Science 2004)

## MISA - MicroSatellite identification tool

This tool allows the identification and localization of perfect microsatellites as well as compound microsatellites which are interrupted by a certain number of bases.

In order to design primers flanking the microsatellite loci, two perl scripts serve as interface modules for the program-to-program data interchange between MISA and the primer modelling software Primer3 (Whitehead Institute). For installation instructions of Primer3 see [http://www-genome.wi.mit.edu/genome\\_software/other/primer3.html](http://www-genome.wi.mit.edu/genome_software/other/primer3.html).

---

[Get MISA](#)

Supplemental perl tools:

[Get est\\_trimmer.pl](#) - perl script useful for pre-preprocessing sequences (invocation without parameters is explaining the syntax)

[Get Primer 3 interface modules](#) - perl scripts allowing the interaction with Primer3

---

Questions and comments: [Thomas Thiel](#)

Last updated: 5/14/02



# 1- Recherche de microsattellites par MISA

Se connecter à l'instance Galaxy locale : [bioinfo.ird.bf:8080/](http://bioinfo.ird.bf:8080/)

**Galaxy** Analyze Data Workflow Shared Data Visualization Admin Help

Tools

STATISTICS/GRAPHICS

- Statistics
- Graph/Display Data
- Multiple regression
- Multivariate Analysis
- Vizualisation

GENOMICS

- Genomics
- Extract Features
- Get Genomic Scores
- ENCODE Tools
- SNiPlay3
- Lift-Over

SOUTHGREEN PROJECTS

- ESTtik
- NGS: Small RNAs
- SignalP Signal peptide and cleavage sites in gram+, gram- and eukaryotic amino acid sequences
- SNiPlay3
- NGS: Picard Tools
- GATK
- BedTools
- MISA: MicroSatellite detection
- Clean fasta header Removes fasta description fields in header
- MISA to GFF3 convert MISA to gvf/gff3
- MISA Primer Design Design primer sets using MISA output
- Detect SSRs using MISA** Detect simple sequence repeats using MISA

Detect SSRs using MISA (version 1.0.0)

fasta Source file:

Min Repeats for Unit Size 1:

Min Repeats for Unit Size 2:

Min Repeats for Unit Size 3:

Min Repeats for Unit Size 4:

Min Repeats for Unit Size 5:

Min Repeats for Unit Size 6:

Interruptions: max\_difference\_between\_2\_SSRs:

Execute

**Importeur les données de séquences depuis la librairie partagée puis lancer le logiciel MISA**

**CITATION**

Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.)

20. Thiel, W. Michalek, R. Varshney and A. Graner

THEORETICAL AND APPLIED GENETICS Volume 106, Number 3, 411-422

DOI: 10.1007/s00122-002-1031-0

**TIP** The primer design tool will work more predictably if fasta header lines are simplified to only include the sequence ID

This tool was installed from a ToolShed, you may be able to find additional information by following this link: <http://toolshed.q2.bx.psu.edu/v>

## 2- Définition de primers

**Galaxy** Analyze Data Workflow Shared Data Visualization Admin Help User

**Tools**

- Multiple Alignments
- Metagenomic analyses
- STATISTICS/GRAPHICS
- Statistics
- Graph/Display Data
- Multiple regression
- Multivariate Analysis
- Vizualisation
- GENOMICS
- Genomics
- Extract Features
- Get Genomic Scores
- ENCODE Tools
- SNIPlay3
- Lift-Over
- SOUTHGREN PROJECTS
- ESTtik
- NGS: Small RNAs
- SignalP Signal peptide and cleavage sites in gram+, gram- and eukaryotic amino acid sequences
- SNIPlay3
- NGS: Picard Tools
- GATK
- BedTools
- MISA: MicroSatellite detection
  - Clean fasta header
    - Removes fasta description fields in header
  - MISA to GFF3 convert MISA to gvf/gff3
  - MISA Primer Design** Design primer sets using MISA output
  - Detect SSRs using MISA
    - Detect simple sequence repeats using MISA

**MISA Primer Design (version 1.0.0)**

**Misa Source file:** 991: Detect SSRs using MISA on data 990

**Fasta Source file:** 990: Chr1.sample.fa

Execute

**i** Design SSR primer sets from MISA output using Primer3

This tool uses helper scripts developed at IPK

<http://pgrc.ipk-qatersleben.de/misa/primer3.html>

**i** This tool was installed from a ToolShed, you may be able to find additional information by following this link: <http://toolshed.g2.bx.psu.edu/view/iohn>

**1** **2** **3** **4** **5** **6** **7** **8**

ID	SSR nr.	SSR type	SSR	size	start	end	FORWARD PRIMER1 (5'-3')
Chr1	1	c	(CT)6tatctctctctttttttcccttcggccctggagcggtgggtaggacgttgacaggtgctgga(GCG)5	93	11185	11277	AAGTTTGGCGTTTATGTGGC
Chr1	2	p2	(CT)7	14	13879	13892	GCCATCCTCCACTACTCCAA
Chr1	3	c	(T)10g(A)10	21	15950	15970	TAAAAGTCGTGCATGTGCGT
Chr1	4	p1	(T)10	10	18622	18631	CTTGCAAGTATTGGTGGCCT
Chr1	5	p1	(A)10	10	20837	20846	AGAATTCCGTGTACCGCAC
Chr1	6	p1	(A)10	10	21731	21740	GCATTGCTCATCATCTTCCC
Chr1	7	p4	(TATC)15	60	23315	23374	AGGGAGCTGGAAGAAGGA
Chr1	8	p1	(T)11	11	26690	26700	GGGGATATGCAAGTGAAGA
Chr1	9	p2	(AT)40	80	27895	27974	CACCTCTATTGCGCGT
Chr1	10	p2	(AG)7	14	29605	29618	CTAACTGGACGGCGATGGTA
Chr1	11	c	(GA)9(G)14	32	29727	29758	GCCAATCCAACATTCGTGTA
Chr1	12	p3	(GGT)6	18	57481	57498	CTTCGGGGTGGTGTAAATTG

# 3- Visualization des primers après génération d'un fichier GFF

Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help User

Tools

Genomics

Extract Features

Get Genomic Scores

ENCODE Tools

SNIPlay3

Lift-Over

SOUTHGREN PROJECTS

ESTtik

NGS: Small RNAs

SignalP Signal peptide and cleavage sites in gram+, gram- and eukaryotic amino acid sequences

SNIPlay3

NGS: Picard Tools

GATK

BedTools

MISA: MicroSatellite detection

Clean fasta header Removes fasta description fields in header

MISA to GFF3 convert MISA to gtf/gff3

MISA Primer Design Design primer sets using MISA output

Detect SSRs using MISA Detect simple sequence repeats using MISA

MISA to GFF3 (version 1.0.0)

Input MISA File: 16: Top Table from DE Seq on data 11  
MISA file from MISA.pl

Execute

TIP

This tool parses MISA columnar output to valid GFF3 format

Example

input

ID	SSR nr.	SSR type	SSR	size	start	end
FW9DLND01C1Y2S	1	p2	(AC)11	22	89	110
FW9DLND01CSIIJ	1	p2	(AC)7	14	226	239
FW9DLND01D0GBK	1	p1	(A)12	12	162	173
FW9DLND01B7MCF	1	p2	(AT)7	14	22	35
FW9DLND01DHCGH	1	p2	(AT)7	14	108	121

output

ID	SSR nr.	SSR type	SSR	size	start	end
FW9DLND01C1Y2S	MISA	microsatellite	89	110	.	.
FW9DLND01CSIIJ	MISA	microsatellite	226	239	.	.
FW9DLND01D0GBK	MISA	microsatellite	162	173	.	.
FW9DLND01B7MCF	MISA	microsatellite	22	35	.	.
FW9DLND01DHCGH	MISA	microsatellite	108	121	.	.

This tool was installed from a ToolShed, you may be able to find additional

## IGV (Integrative Genome Viewer)



Charger 3 fichiers:

- Fichier Fasta de la région génomique
- Annotation GFF des gènes de cette région
- Fichier GFF de primers

