

Galaxy: ressource pour le calcul en bioinformatique

Plateaux Bioinformatiques du CIRAD, et de l'IRD
pour la plateforme Southgreen

Alexis Dereeper, Jean-François Dufayard, Cédric Farcy, Ayité KOUGBEADJO, Pierre Larmande, Christine Tranchant

avec l'aide des formateurs précédents: Félix Homa, Stéphanie Pointet, Marilyne Summo

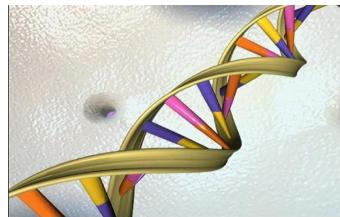


La bioinformatique sans Galaxy

La bioinformatique dans des équipes de biologie



Problématique relevant de la bioinformatique

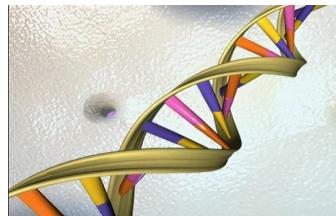


```
ACGGTCTTAA  
AGATGTTCTA  
TAGGTAAAGAAC  
AGCTCATTCAGCTT  
TTGGATGCTACTG  
TGGGATACCCCTGG  
CGTCACATTCAAACGG  
CGTCACTTAAAGGCC  
CGAGCGTCCGGTTAA  
CTTACTGACGCAGA  
TAAAACGTTCTGGCG  
AAAAACGTTCTGGTA  
CTCGTCTTGGTATGTA  
GCTCGTAAATATTCAA  
ATGGTCGCTTATTAC  
CTCCGCTTCTGCA  
CTCATCGTCAC  
TGTTAAC  
TAN
```

La bioinformatique dans des équipes de biologie



Problématique relevant de la bioinformatique



A grid of DNA sequence data, showing multiple lines of text representing nucleotide sequences. The sequences include various bases (A, T, C, G) and some gaps or stop signs.



Je dispose d'un bioinformaticien sur place



La bioinformatique dans des équipes de biologie



Problématique relevant de la bioinformatique



A grid of DNA sequence data, showing multiple lines of nucleotide sequences (A, T, C, G) representing a dataset for bioinformatics analysis.



Je dispose d'un bioinformaticien sur place

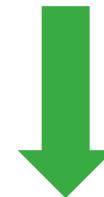
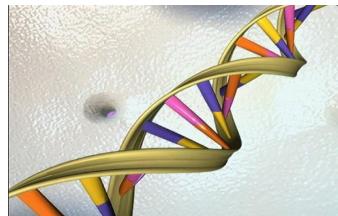


- Il n'est pas omniscient.
- Il n'est pas omnipotent.
- Il ne dispose que d'un temps limité.
- Il est dépendant d'une structure pour le calcul intensif.

La bioinformatique dans des équipes de biologie



Problématique relevant de la bioinformatique



Je dispose d'une plateforme



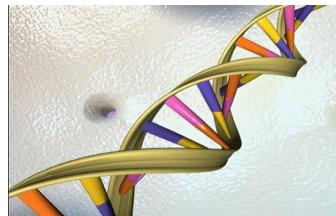
Je dispose d'un bioinformaticien sur place

- Il n'est pas omniscient.
- Il n'est pas omnipotent.
- Il ne dispose que d'un temps limité.
- Il est dépendant d'une structure pour le calcul intensif.

La bioinformatique dans des équipes de biologie



Problématique relevant de la bioinformatique



Je dispose d'une plateforme

- Identifier le bon interlocuteur est délicat.
- Le temps de réaction peut varier.



Je dispose d'un bioinformaticien sur place

- Il n'est pas omniscient.
- Il n'est pas omnipotent.
- Il ne dispose que d'un temps limité.
- Il est dépendant d'une structure pour le calcul intensif.

Parfois, l'analyse est "simple", et consiste à employer des outils connus...

La technique du "Je Le Fais Moi-Même"

Dans cette situation, la solution la plus fréquemment employée est le

JLF2M: Je le fais moi-même !

La technique du "Je Le Fais Moi-Même"

Dans cette situation, la solution la plus fréquemment employée est le

JLF2M: Je le fais moi-même !

```
simmons@earth:~/work/xterm/orig/xterm-XF86-4.3.0$ ls
256cols.h      error.h      misc.c      Tekparse.h    VTPrsTbl.c
256cols.pl     fontutils.c   mdirs.sh    Tekproc.c    vttests
88cols.h       fontutils.h   os2main.c   TekPrsTbl.c  wcidth.h
88cols.pl      Imakefile    plink.sh   tektests
AAA_README_VMS.txt input.c    precompose.c termcap    xcharmouse.h
acllocal.m4    INSTALL     precompose.h terminfo   xstrings.c
button.c       install.sh   print.c    Tests       xstrings.h
charclass.c    keysym2ucs.c proto.h    testxmc.c  XTerm.ad
charclass.h    keysym2ucs.h ptydata.c   trace.c    xterm_axp.opt
charproc.c     link_axp.com ptx.h     trace.h    xtermcfg.hin
charsets.c    main.c      README     unicode   XTerm-col.ad
config.guess   make.com    resize.c   uxterm    xterm.dat
config.sub     Makefile    resize.man UXTerm.ad xterm.h
configure      Makefile.in screen.c   version.h xterm_io.h
configure.in   Makefile.in scrollba.c vms.c     xterm_man
cursor.c       MANIFEST    tabs.c    VTparse.def xutf8.c
data.c         menu.c     sinstall.sh vms.h     xutf8.h
data.h         menu.h     tabs.c    VTparse.def
doublechr.c   minstall.sh Tekparse.def
simmons@earth xterm-XF86-4.3.0$
```

Une invite de commandes



Un cluster de calcul

La technique du "Je Le Fais Moi-Même"

```
simmons@earth:~/work/xterm/orig/xterm-XF86-4.3.0]$ ls
256colres.h      error.h      misc.c      Tekparse.h    VTPRstTbl.c
256colres.pl     fontutils.c  midirs.sh   Tekproc.c    vttests
88colres.h       fontutils.h  os2main.c  TekPrstTbl.c  wcoutwidth.c
88colres.pl     Iwakefile   plink.sh    tektests   wcoutwidth.h
AAA_README_VMS.txt input.c    precompose.c termcap    xcharmouse.h
aclocal.m4        INSTALL    precompose.h terminfo   xstrings.c
button.c         install.sh  print.c    Tests      xstrings.h
charclass.c      keysym2ucs.c proto.h    testxmc.c  XTerm.ad
charclass.h      keysym2ucs.h ptydata.c  trace.c    xterm_axp.opt
charproc.c       link_axp.com ptyx.h     trace.h    xtermcfg.hin
charsets.c       main.c     README     unicode   XTerm-col.ad
config.guess     main.h     README.os390 util.c    xterm.dat
config.sub       make.com   resize.c   xterm      xterm.h
configure        Makefile   resize.man UXTerm.ad xterm_io.h
configure.in     Makefile.in screen.c   version.h xterm_log.html
cursor.c         MANIFEST   scrollbar.c vms.c    xterm.man
data.c          menu.c    sinstall.sh vms.h    xutf8.c
data.h          menu.h    tabs.c    VTParse.def xutf8.h
doublechr.c     minstall.sh Tekparse.def VTParse.h
[simmons@earth xterm-XF86-4.3.0]$
```

Communication



- Apprendre les commandes Linux de base.
- Apprendre les commandes spécifiques à la communication avec un cluster de calcul.

La technique du "Je Le Fais Moi-Même"



Utiliser les logiciels en ligne de commande, les manuels en ligne ou en interface texte.

Command line parameters

All program parameters can be entered in the command line. The first parameter must always be the name of the alignment file or the pathnames file. If this is the only given parameter, the menu of the program is activated. For example:

Gblocks nad3.pir

The next parameters are entered according to the letter of the corresponding menu item. They can be entered in any order. The list of all parameters is:

PARAMETER NAME	MEANING (Default)	ALLOWED VALUES
(None)	Filename (No default)	Alignment or pathnames file
-t=	Type Of Sequence (Protein, DNA, Codons)	p, d, c
-b1=	Minimum Number Of Sequences For A Conserved Position (50% of the number of sequences + 1)	Any integer bigger than half the number of sequences and smaller or equal than the total number of sequences
-b2=	Minimum Number Of Sequences For A Flank Position (85% of the number of sequences)	Any integer equal or bigger than Minimum Number Of Sequences For A Conserved Position
	Maximum Number Of Contiguous	

```
dufayard@marmadais:/home/dufayard/tmp
dufayard@marmada... dufayard@marmada... dufayard@marmada... dufayard@marmada... Local
Execute Bookmarks

-dt species_threshold
    The support threshold for gene tree branch collapse (optional, default 80.0)

-pt polymorphism_threshold
    The length threshold for species tree branch collapse (optional, default 10.0)

-k k_level
    The k-level of the subtree-neighbor measure (optional, default 2)

-idupw i_duplication_weight
    The weight of intersection duplication in functional orthology scoring (0.0 for maximum weight, 1.0 for no weight, optional, default 0.90)

-tdupw t_duplication_weight
    The weight of topological duplication in functional orthology scoring (0.0 for maximum weight, 1.0 for no weight, optional, default 0.95)

-specw speciation_weight
    The weight of speciation in functional orthology scoring (0.0 for maximum weight, 1.0 for no weight, optional, default 0.99)

-ultraw ultraparalogy_weight
    The weight of an ultraparalogy node in functional orthology scoring (0.0 for maximum weight, 1.0 for no weight, optional, default 0.99)

-distw distance_weight
    The weight of evolutionary distance in functional orthology scoring (0.0 for maximum weight, 1.0 for no weight, optional, default 0.10)

[root@marmadais tmp]#
```

La technique du "Je Le Fais Moi-Même"



- Assumer une veille technologique.
- Apprendre à gérer la compatibilité entre formats d'entrées et de sorties.
- Apprendre à faire voyager les données entre cluster de calcul et serveurs externes en ligne.
- Apprendre un langage de scripts.

```
RapGreen.java - /Volumes/dufayard/RAP_green/rap-green/trunk/rappgreen (rapgreen)
Nouveau Ouvrir Enregistrer Fermer
Recherche dynamique Recherche avancée Info
BuildSpeciesTree.java 216
BuildSpeciesTree.java 217
BuildSpeciesTree.java 218
BuildSpeciesTree.java 219
BuildSpeciesTree.java 220
BuildSpeciesTree.java 221
BuildSpeciesTree.java 222
BuildSpeciesTree.java 223
BuildSpeciesTree.java 224
BuildSpeciesTree.java 225
BuildSpeciesTree.java 226
BuildSpeciesTree.java 227
BuildSpeciesTree.java 228
BuildSpeciesTree.java 229
BuildSpeciesTree.java 230
BuildSpeciesTree.java 231
BuildSpeciesTree.java 232
BuildSpeciesTree.java 233
BuildSpeciesTree.java 234
BuildSpeciesTree.java 235
BuildSpeciesTree.java 236
BuildSpeciesTree.java 237
BuildSpeciesTree.java 238
BuildSpeciesTree.java 239
BuildSpeciesTree.java 240
BuildSpeciesTree.java 241
BuildSpeciesTree.java 242
BuildSpeciesTree.java 243
BuildSpeciesTree.java 244
BuildSpeciesTree.java 245
BuildSpeciesTree.java 246
BuildSpeciesTree.java 247
BuildSpeciesTree.java 248
BuildSpeciesTree.java 249
BuildSpeciesTree.java 250
BuildSpeciesTree.java 251
BuildSpeciesTree.java 252
BuildSpeciesTree.java 253
BuildSpeciesTree.java 254
BuildSpeciesTree.java 255
BuildSpeciesTree.java 256
BuildSpeciesTree.java 257
BuildSpeciesTree.java 258
BuildSpeciesTree.java 259
BuildSpeciesTree.java 260
BuildSpeciesTree.java 261
BuildSpeciesTree.java 262
BuildSpeciesTree.java 263
BuildSpeciesTree.java 264
Arbre798_rerooted.tree
Arbre798.tree
Arbre798_annotated.tree
Enregistrement: 11 juin 2012 16:55 - Longueur: 14 271 - Sélection: 13 - Encodage: Unicode (UTF-8)
```

The screenshot shows the Castresana Lab website with a focus on the Gblocks Server. The page includes a brief description of the program's purpose, a text input field for alignment paste, and various selection options for phylogenetic analysis.

Castresana Lab
Animal Phylogeny and Systematics Program
[Home](#) | [People](#) | [Research](#) | [Publications](#) | [Software](#)

Gblocks Server
Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis

About the Gblocks Server
Version 0.91b, January 2002
Copyright © Jose Castresana

Gblocks eliminates poorly aligned positions and divergent regions of a DNA or protein alignment so that it becomes more suitable for phylogenetic analysis. This server implements the most important features of the Gblocks program to make its use as simple as possible without loosing the functionality that it is necessary in most of the cases. Other options can be changed in the stand-alone program. You can see here an [example output file](#) showing the Gblocks results from a protein alignment. Further information can be found in the [online documentation](#). Please see the Gblocks page for citations.

Gblocks Server
Paste an alignment in NBRF/PIR or FASTA format:

Or upload an alignment file:
 Choisir un fichier Aucun fichier choisi

Type of sequence:
 DNA Protein Codons

Options for a less stringent selection:
 Allow smaller final blocks
 Allow gap positions within the final blocks
 Allow less strict flanking positions

Options for a more stringent selection:
 Do not allow many contiguous nonconserved positions

Get Blocks Clear



Galaxy, rôle et utilisation ?

Galaxy

Galaxy se place directement dans le JLFM2

```
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  ucwidth.c
88colres.pl     Imakefile   plink.sh   tektests   ucwidth.h
AAA_README_VMS.txt input.c    precompose.c termcap   xcharmouse.h
aclocal.m4        INSTALL    precompose.h terminfo  xstrings.c
button.c         install.sh  print.c    Tests      xstrings.h
charclass.c      keysym2ucs.c proto.h    testxmc.c  XTerm.ad
charclass.h      keysym2ucs.h ptydata.c  trace.c    xterm_axp.opt
charproc.c       link_axp.com ptxx.h    trace.h    xtermcfg.hin
charsets.c       main.c     README    unicode   XTerm-col.ad
config.guess     main.h     README.os390 util.c    xterm_dat
config.sub       make.com   resize.c   uTerm     xterm.h
configure        Makefile   resize.man UXTerm.ad xterm_io.h
configure.in     Makefile.in screen.c   version.h xterm_log.html
cursor.c         MANIFEST   scrollbar.c vms.c    xterm_man
data.c           menu.c    sinstall.sh vms.h    xutf8.c
data.h           menu.h    tabs.c    VTparse.def xutf8.h
doublechr.c     minstall.sh Tekparse.def VTparse.h
[simmons@earth xterm-XF86-4.3.0]$
```

Communication



Communication

Galaxy

Galaxy se place directement dans le JLFM2

The screenshot shows the Galaxy web interface integrated with the South Green bioinformatics platform. The top navigation bar includes links for Analyze Data, Workflow, Shared Data, Admin, Help, and User, along with the South Green logo.

The main content area features the South Green bioinformatics platform branding, including the logo and the text "Welcome to GALAXY ... at your disposal as part of the services provided by SouthGreen".

A sidebar on the left lists various tools and resources, such as Tools, Recently Used, Get Data, Send Data, and a comprehensive list of Tools including Convert Formats, Evolution, ESTtik, Filter and Sort, Gene/Protein prediction, SAT, NGS: Quality Control, NGS: Mapping, NGS: SAM/BAM Manipulations, NGS: SNP Detection, Protein Structures, Sequence comparisons, and UNTESTED TOOLS like Text Manipulation, Filter and Sort, Join, Subtract and Group, Convert Formats, Extract Features, and Fetch Sequences.

The right side displays a History panel listing recent jobs:

Job ID	Description	Status	Size
20: PhyML on data 18	PhyML on data 18	Completed	4.0 Mb
19: PhyML on data 18	PhyML on data 18	Completed	4.0 Mb
18: Fasta2Phylip on data 16	Fasta2Phylip on data 16	Completed	4.0 Mb
17: Gblocks on data 15	Gblocks on data 15	Completed	4.0 Mb
16: Gblocks on data 15	Gblocks on data 15	Completed	4.0 Mb
15: MAFFT on data 14	MAFFT on data 14	Completed	4.0 Mb
14: new.fasta	new.fasta	Completed	4.0 Mb

Two callout boxes provide additional information:

- A yellow dashed box: "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."
- A green dashed box: "Requests for making new tools available shall be addressed to admin.bioinfo@cirad.fr"

At the bottom, a footer note states: "The GALAXY project is supported in part by NSF, NHGRI, and the Huck Institutes of the Life Sciences."

Galaxy

Galaxy se place directement dans le JLFM2

Accès direct au cluster

Centralisation et partage des données

The screenshot shows the Galaxy web interface. At the top, there is a navigation bar with tabs: Analyze Data, Workflow, Shared Data, Admin, Help, and User. The User tab is currently active, showing the South Green bioinformatics platform logo. Below the navigation bar, there is a sidebar on the left containing links for Tools (search tools, Recently Used, Get Data, Send Data), TOOLS (Convert Formats, Evolution, ESTtik), and a section for NGS: SNP Detection, Protein Structures, Sequence comparisons. There is also a section for UNTESTED TOOLS with links to Text Manipulation, Filter and Sort, Join, Subtract and Group, Convert Formats, Extract Features, and Fetch Sequences. The main content area features the South Green logo and the text "Centralisation et partage des démarches". It also includes a "Welcome to GALAXY" message and a note about tool availability: "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." Another note states: "Requests for making new tools available shall be addressed to admin.bioinfo@cirad.fr". At the bottom, it says "The GALAXY project is supported in part by NSF, NHGRI, and the Huck Institutes of the Life Sciences." To the right, there is a vertical sidebar titled "History" listing several recent activities: "PhyML on data 18", "PhyML on data 18", "18: Fasta2Phylip on data 16", "17: Gblocks on data 15", "16: Gblocks on data 15", "15: MAFFT on data 14", and "14: new.fasta".

Interface et compte utilisateur indépendants du poste client

Utilisable depuis n'importe quel poste connecté à internet

Galaxy - Vue générale

The screenshot shows the Galaxy bioinformatics platform interface. At the top, there is a navigation bar with tabs: Analyze Data, Workflow, Shared Data, Admin, Help, and User. The User tab is highlighted. On the left, there is a sidebar titled "Boîte à outils" (Tool Box) which lists various tools categorized under Tools, Recently Used, Get Data, Send Data, TOOLS, INTESTED TOOLS, and SouthGreen. A red circle highlights the "Tools" section of the sidebar. The main content area displays the "South Green bioinformatics platform" logo and the "Welcome to GALAXY" message. It also includes two informational boxes: one about tool availability and another about tool requests. At the bottom, it mentions support from NSF, NHGRI, and the Huck Institutes of the Life Sciences. On the right, there is a "History" panel showing a list of recent jobs, each with a preview icon, a name, and a size (e.g., 4.0 Mb). A red circle highlights the first job in the history list. A red box at the bottom right is labeled "Données".

Boîte à outils

Analyze Data Workflow Shared Data Admin Help User

Tools

search tools

Recently Used

Get Data

Send Data

TOOLS

Convert Formats

Evolution

ESTtik

Filter and Sort

Gene/Protein prediction

SAT

NGS: Quality Control

NGS: Mapping

NGS: SAM/BAM Manipulations

NGS: SNP Detection

Protein Structures

Sequence comparisons

INTESTED TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

Extract Features

cirad IRD Institut de recherche pour le développement Bioversity INRA SupAgro

Welcome to GALAXY

... at your disposal as part of the services provided by SouthGreen

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.

Requests for making new tools available shall be addressed to admin.bioinfo@cirad.fr

The GALAXY project is supported in part by NSF, NHGRI, and the Huck Institutes of the Life Sciences.

History

Options

Unnamed history

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

4.0 Mb

Données

Galaxy - Vue générale

Chaque programme peut être configuré précisément

Configuration d'un programme

Tools Options ▾

- conservation index of proteins
- [Blast+](#) More than a Blast...
- [Exonerate](#) a generic tool for sequence alignment
- [Gblocks](#), an alignment curator
- [Genome Threader](#) Gene Prediction Software
- [HMMBUILD](#), profile HMM construction from multiple sequence alignments
- [HMMCALIBRATE](#), calibrate HMM search statistics
- [MAFFT](#), a progressive multiple alignment builder
- [MUSCLE](#), a progressive multiple alignment builder
- [BlastClust](#), a blast-based clustering method
- [UClust](#), a seed-based clustering method
- [BLAST+ blastn \(MC\)](#) Search nucleotide database with nucleotide query sequence(s)
- [BLAST+ blastp \(MC\)](#) Search protein database with protein query sequence(s)
- [BLAST+ blastx \(MC\)](#) Search protein database with

Analyze Data Workflow Clusters Help User

UClust (version 1.0.0)

Source file: 20: PhyML on data 18

Fasta sequence file

Minimum similarity ratio: 0.50

Ratio between 0 and 1

Execute

Program encapsulated in Galaxy by Southgreen

UClustClust version 3.0.617, 2010

Please cite:

"Search and clustering orders of magnitude faster than BLAST", Edgar RC, Bioinformatics 2010 doi: 10.1093/bioinformatics/btq461.

Overview

UBLAST and USEARCH are algorithms designed to enable high-throughput, sensitive search of very large sequence databases. UBLAST searches for local alignments, USEARCH for global alignments. UBLAST and USEARCH are orders of magnitude faster than BLAST for some applications. UCLUST is a clustering algorithm that uses USEARCH as a subroutine to achieve exceptional high speed and sensitivity. The algorithms are described in this paper:

For further informations, please visite the [UClust website](#).

History Options ▾

- Unnamed history 4.0 Mb
- 20: PhyML on data 18
- 19: PhyML on data 18
- 18: Fasta2Phyphil on data 16
- 17: Gblocks on data 15
- 16: Gblocks on data 15
- 15: MAFFT on data 14
- 14: new.fasta

Galaxy - accès aux données

Les données peuvent être chargées depuis votre ordinateur local

Pour importer depuis votre ordinateur local

Pour importer des données depuis une autre page web ou par copier/coller

Explication sur les différents formats

Galaxy

Analyze Data Workflow Shared Data Admin Help User

South Green bioinformatics platform

Tools Options

search tools

Recently Used

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 Test server
- CBI Rice Mart rice mart
- GrameneMart Central server
- modENCODE fly server
- Flymine server
- Flymine test server
- modENCODE modMine server
- Ratmine server
- modENCODE worm server
- Wormbase server
- Wormbase test server
- EuPathDB server
- EncodeDB at NHGRI
- EpiGRAPH server
- EpiGRAPH test server
- HbVar Human Hemoglobin Variants and Thalassemias

Upload File (version 1.1.3)

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

File: Choisissez un fichier Aucun fichier choisi

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

URL/Text:

Here you may specify a list of URLs (one per line) or paste text here.

Convert spaces to tabs: Yes

Use this option if you are entering intervals by hand.

Genome: Click to Search or Select

Execute

Auto-detect

The system will attempt to detect Axt, Fasta, Fastqsolexa, Gff, Gff3, Html, Lax, Maf, Tabular, Wigs, Bed and Interval (Bed with headers) formats. If your file is not detected properly as one of these formats, it may still be uploaded successfully. If you experience some format problems (e.g., different number of columns on each line), you can manually select the file format you think it should be. You can also upload compressed files (e.g., zip, tar, gzip). Note that compressed files must be decompressed before they are processed.

Ab1

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

History Options

- Unnamed history 4.0 Mb
- 20: PhyML on data_18
- 19: PhyML on data_18
- 18: PhyML on data_18
- 17: PhyML on data_18
- 16: PhyML on data_18
- 15: PhyML on data_18
- 14: new.fasta

Galaxy - accès aux données

Les données peuvent être chargées depuis des bibliothèques partagées

The screenshot shows the Galaxy web interface. At the top, there is a navigation bar with links for 'Analyze Data', 'Workflow', 'Shared Data' (which is highlighted with a red circle), 'Admin', 'Help', and 'User'. To the right of the navigation bar is the 'South Green bioinformatics platform' logo. Below the navigation bar, the title 'Data Library "Formation"' is displayed. The main content area shows a list of datasets in the 'Formation' library. One specific dataset, 'exemple.fasta', is highlighted with a red circle and labeled 'Données publiques' (Public Data) in a red box. The dataset details are as follows:

Name	Message	Data type	Date uploaded	File size
Annotation_gene				
Annotation_transcriptome				
Debuts dans Galaxy	Données pour l'initiation à Galaxy			
exemple.fasta	Exemple de fichier fasta utilisé dans la démonstration de Galaxy	fasta	2012-02-06	15.2 Kb
Pre-processing and Ma				
PreProcessing and Map				
SNP	SNP			
TD Annotation 2012				

At the bottom left, there is a note: 'For selected datasets: Import to current history Go'. Below the table, there are two tips:

- 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 Galaxy bioinformatics platform interface. On the left, there's a sidebar titled "History" with a list of recent histories: "Click to rename history" (Unnamed history, 4.0 Mb), "20: PhyML on data 18", "19: PhyML on data 18", "18: Fasta2Phyphil on data 16", "17: Gblocks on data 15", "16: Gblocks on data 15", "15: MAFFT on data 14", and "14: new.fasta". A red box highlights the "Historique courant" (Current History) entry. An "Options" dropdown menu is open over the list, containing the following items: "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", and "Import from File".

Saved Histories

search history names and tags Advanced Search

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

Listes des historiques personnels, et publiés

Published Histories

search name, annotation, owner, and tag Advanced Search

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

Galaxy - Historiques des analyses

The screenshot shows the Galaxy web interface for managing histories. On the left, a sidebar menu is open, showing options like 'History Lists', 'Saved Histories' (highlighted with a red box), and 'Delete Permanently' (also highlighted with a red box). A red arrow points from the 'Delete Permanently' option to the 'Delete' link in the main content area. The main content area is titled 'Saved Histories' and contains a table of analysis histories. A red box highlights the 'Advanced Search' link at the top of the table header. Another red box highlights the 'Selectionnez les historiques à supprimer' (Select histories to delete) message above the table. A red arrow points from the 'Delete' link in the sidebar to the 'Delete' link in the table header. Two red speech bubbles are present: one pointing to the 'Delete' link in the table header with the text 'Suppression temporaire' (Temporary deletion), and another pointing to the 'Delete Permanently' button with the text 'Suppression définitive' (Definitive deletion).

Name	Datasets	Tags	Sharing	Size on Disk	Created	Last Updated	Status
		0 Tags		0 bytes	less than a minute ago	less than a minute ago	current history
		0 Tags		0 bytes	3 days ago	~ 4 hours ago	
IGC3_combine	4	0 Tags		590.0 Mb	~ 6 hours ago	~ 5 hours ago	
Iq010 Clean Data	4	1	0 Tags	38.6 Gb	Jun 04, 2012	3 days ago	
Iq09 Clean Data	4	1	0 Tags	38.0 Gb	Jun 04, 2012	3 days ago	
Iq08 Clean Data	4	1	0 Tags	28.9 Gb	Jun 04, 2012	3 days ago	
Iq07 Clean Data	4	1	0 Tags	25.4 Gb	Jun 04, 2012	3 days ago	
Iq06 Clean Data	5	0 Tags		26.9 Gb	Jun 04, 2012	3 days ago	
Iq05 Clean Data	4	1	0 Tags	31.9 Gb	Jun 04, 2012	3 days ago	
Iq04 Clean Data	4	1	0 Tags	35.2 Gb	Jun 04, 2012	3 days ago	
Iq03 Clean Data	4	1	0 Tags	27.7 Gb	Jun 04, 2012	3 days ago	
Iq02 Clean Data	4	1	0 Tags	28.6 Gb	Jun 04, 2012	3 days ago	
Iq01 Clean Data	4	1	0 Tags	30.3 Gb	May 31, 2012	3 days ago	
arcad test data	3	1	0 Tags	2.6 Gb	Apr 24, 2012	Jun 11, 2012	
Cleaning_NGS_felix	4	0 Tags		0 bytes	May 07, 2012	May 07, 2012	
galaxy_test	17	0 Tags		0 bytes	May 04, 2012	May 07, 2012	

Galaxy - Historique des analyses

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

History

Click to rename history

Clean FASTQ 37.6 MB

15: Concatenate datasets on data 13 and data 14

5: Cutadapt on data 2 (Report)

3: Cutadapt on data 1 (Report)

Historique courant

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

The screenshot shows the Galaxy web interface. At the top, there's a navigation bar with 'Tools' and 'Options' menus. Under 'Tools', there are links for 'search tools', 'Recently Used', 'Get Data', 'Send Data', and 'TOOLS'. A red circle highlights the 'TOOLS' link. Below this, a red box contains the text 'Outils testés et maintenus par Southgreen'. The main content area lists several tool categories with their sub-tools:

- Convert Formats**
- Evolution**
 - FastME , a distance tree builder (NJ, BioNJ, NNI or SPR optimized)
 - PhyML , a ML tree builder
 - RAP-Green , a species/gene tree reconciler
 - SDI_r , a species/gene tree reconciler
 - Mutate Codons with SNPs
- ESTtik**
- 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** (highlighted with a red circle)
 - Text Manipulation
 - Filter and Sort

At the bottom, the SouthGreen logo is visible with the text 'bioinformatics platform'.

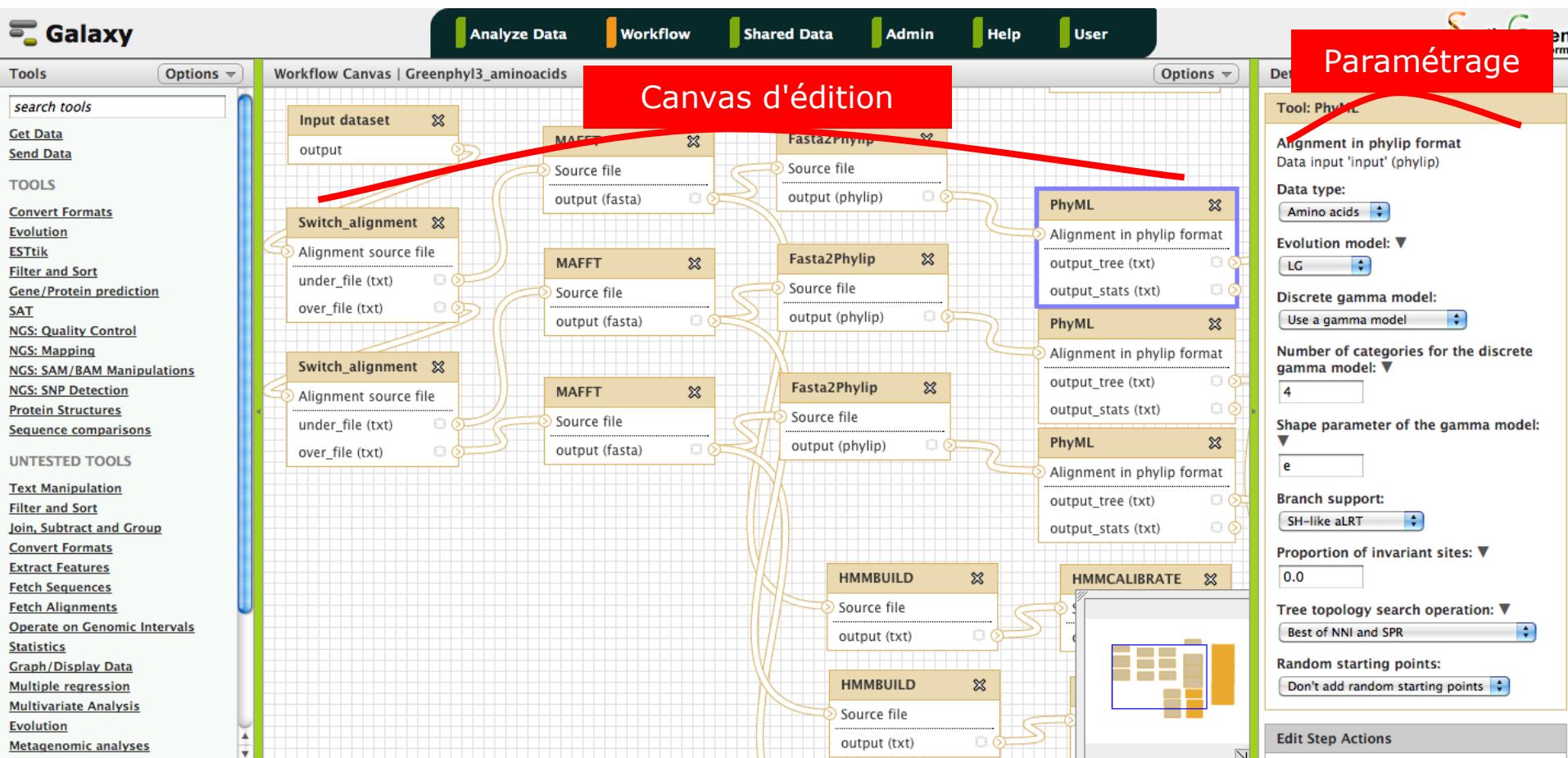
Les outils sont (trop ?) nombreux,
et classés par catégories

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

Outils installés sur le Galaxy original

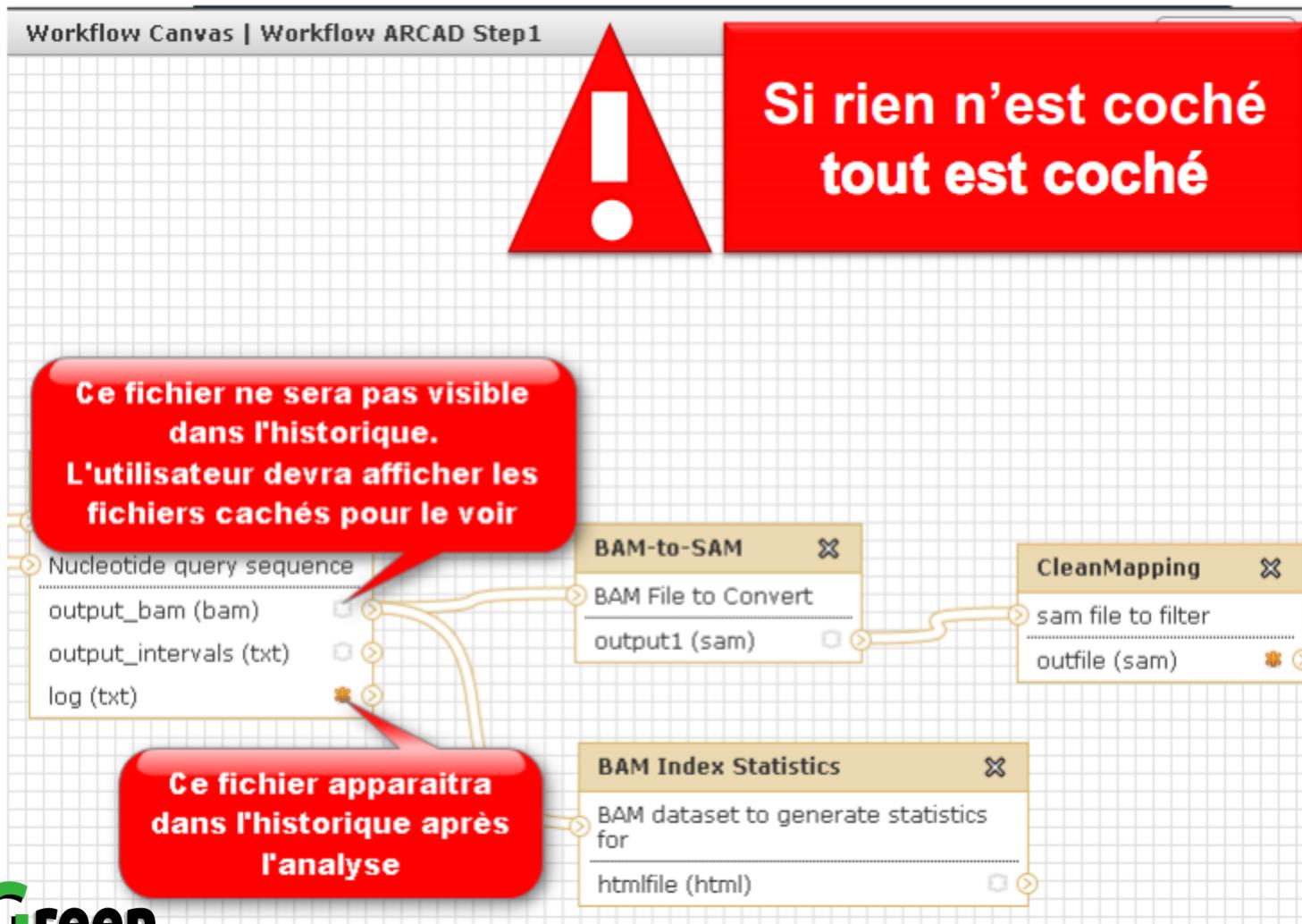
Galaxy - Workflows

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



Galaxy - Workflows

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



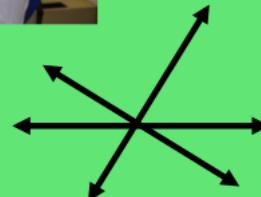


Galaxy au sein de Southgreen

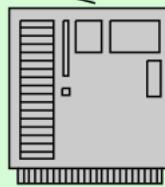
Plateau, plateforme, équipe de recherche



IRD



Equipe Intégration Des Données, UMR AGAP



Plateau : cluster,
serveurs, réseaux,
locaux, etc.



IRD



Galaxy, sa place dans la plateforme

Utilisateur averti

Données à analyser

The screenshot shows the Galaxy web interface. On the left, a sidebar lists various tools categorized under 'TOOLS' and 'UNTESTED TOOLS'. The main area displays a 'Welcome to GALAXY' page from the SouthGreen bioinformatics platform. A 'History' panel on the right shows a list of recent analyses, each with a preview icon and a delete button.

Utilisateur expert

```
simmons@earth xterm~$ 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  wcidth.c
88colres.pl    Imakefile    plink.sh   tektests    wcidth.h
AAA_README_VMS.txt input.c    precompose.c termcap     xcharmouse.h
aclocal.m4       INSTALL     precompose.h terminfo   xstrings.c
button.c        install.sh  print.c    Tests       xstrings.h
charclass.c     keysym2ucs.c proto.h    testxmc.c  XTerm.ad
charclass.h     keysym2ucs.h ptydata.c  trace.c    xterm_apx.opt
charproc.c      link_apx.com main.c    README.os390 util.c    xterm_apx_opt
charset.c       make.com    resize.c   trace.h    xtermcfg.hin
config.guess    Makefile    resize.h   unicode    XTerm-col.ad
config.sub      Makefile.in screen.c  vms.c     xterm_dat
configure       Makefile.in scrollba.c vms.h     xterm.h
configure.in    MANIFEST    sinstall.sh vms.h     xterm_io.h
cursor.c        menu.c     tabs.c    VTparse.def xterm_log.html
data.c          menu.h
data.h          menu.h
doublechr.c    minstall.sh Tekparse.def VTparse.h xterm.man
doublechr.c    minstall.sh Tekparse.def VTparse.h xutf8.c
doublechr.c    minstall.sh Tekparse.def VTparse.h xutf8.h
[simmons@earth xterm~$
```

Reproductibilité des analyses

Données analysées et explorables

The screenshot shows the South Green bioinformatics platform homepage. It features a navigation bar with links to Home, Tools, Databases, Trainings, Downloads, QuickR, and Alfresco. The main content area includes sections for 'OryGenesDB', 'Oryza Tag Line', 'EURIGEN', 'GNP annot', 'SAT', 'TropGENE', 'SNIPlay', 'GNDIVERSITY', and a 'Galaxy' section. A sidebar on the left provides links to various platform components and resources.



Galaxy, son évolution et sa maintenance

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...).

Séances de travail "bioanalystes"

Demi-journées sur différentes thématiques:

- GBS
- RNA-seq
- Bases de données
- ...



Galaxy, démonstration