

# Galaxy: ressource pour le calcul en bioinformatique

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

---

Alexis Dereeper, Jean-François Dufayard, Marilyne Summo, Gautier Sarah

avec l'aide des formateurs précédents: Félix Homa, Stéphanie Pointet, Cédric Farcy



---

# La bioinformatique sans Galaxy

# La bioinformatique dans des équipes de biologie



Problématique relevant de la bioinformatique



```
ACGGTCTAA  
AGATGTTCTA  
ATAGGTAGAAC  
AGCTCATCAGCTT  
TTGGATGCTACTG  
TGTGGATACCCCTGG  
CGTCACATTCAAACGG  
CGTCACTTCAAGGG  
CGAGCGTCCGGTTAA  
CTTACTGACGCAGA  
TCTAAACGTTCTGG  
AAAAAAACGTTGGTA  
CTCGTCTTGGTATGTA  
GCTCGTAAATATTCAA  
ATGGTCGCTTATTACC  
CTCTCCGTCCTTCTC  
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 consist of four characters: A (Adenine), T (Thymine), C (Cytosine), and G (Guanine).

```
ACGGTCTTAA  
AGATTTCTCA  
ATAGGTAAAGAAC  
TGCTCATTCAGCTT  
TTGGATGCTACTGC  
TTGGGATACCCCTCGCT  
TGTGGATTCACGGCT  
CGTCACATTCAAACGGCT  
CGTCACTTCAAGGGCT  
CGAGCGTCCGGTTAAAGGCCGTA  
CTTACTGACGAGAAANACGCT  
CGAGCGTCCGGTTAAAGGCCGTA  
TCTTAACTGCTTGTAGTGTC  
AAAAAAAGCTTGTAGTGTC  
TCTTAACTGCTTGTAGTGTC  
CTCGTCTTGTAGTGTC  
GCTCGTAAATATTCAAACCT  
GCTCGTCTTGTAGTGTC  
ATGGTCGCTTATTACCA  
CTCTCCGTCCTTCTC  
CTCATCGTCAC  
TGTTAAC  
TAN
```



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

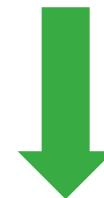


- Il n'est pas omniscient.
- Il n'est pas omnipotent.
- Il ne dispose que d'un temps limité.
- Il est dépendant d'une structure pour le calcul intensif.

# 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  wowitzh.c
88cols.pl      Imakefile    plink.sh   tektests   wowitzh.h
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   main.h       README.os390 util.c    xterm.dat
config.sub     make.com    resize.c   uxterm    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 tabs.c   VTparse_def xutf8.c
data.h         menu.h     tabs.c    VTparse.h xutf8.h
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  mdirs.sh    Tekproc.c    vttests
88colres.h       fontutils.h  os2main.c  TekPrstTbl.c  wcoutwidth.c
88colres.pl     Imakefile    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      resize.c  uterm     xterm.dat
config.sub       make.com    screen.c  version.h xterm.h
configure        Makefile    resize.man UXTerm.ad xterm_io.h
configure.in     Makefile.in scrollbar.c vms.c    xterm.log.html
cursor.c         MANIFEST    menu.c    vms.h     xterm.man
data.c          menu.h      sinstall.sh tabs.c    utf8.c
data.h          menu.h      Tekparse.def VTParse.h  utf8.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	

A screenshot of a terminal window titled 'dufayard@marmadais:/home/dufayard/tmp'. The window contains a list of command line parameters for the 'Gblocks' program. The parameters listed are: -st species\_threshold, -pt polymorphism\_threshold, -k k\_level, -idupw i\_duplication\_weight, -tdupw t\_duplication\_weight, -specw speciation\_weight, -ultraw ultraparalogy\_weight, and -distw distance\_weight. Each parameter is followed by its description and default value. The terminal prompt at the bottom is '[root@marmadais tmp]#'.

```
dufayard@marmadais:/home/dufayard/tmp
dufayard@marmadais:/home/dufayard/tmp dufayard@marmadais:/home/dufayard/tmp dufayard@marmadais:/home/dufayard/tmp Local
The support threshold for gene tree branch collapse (optional, default 80.0)

-st species_threshold
    The length threshold for species tree branch collapse (optional, default 10.0)

-pt polymorphism_threshold
    The length depth threshold to deduce to polymorphism, allelism ... (optional, default 0.05)

-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]#
```

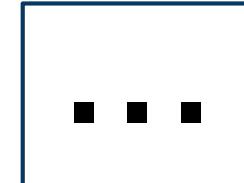
# La technique du "Je Le Fais Moi-Même"



Interopérabilité



Interopérabilité



Construire une analyse complète

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/rappreen (rapgreen)
...
BuildSpeciesTree.java 216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
```

Castrasana 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 Castrasana

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 blocks selected 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:  
 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



---

# 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 ptx.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 SouthGreen bioinformatics platform. The top navigation bar includes links for Analyze Data, Workflow, Shared Data, Visualization, Admin, Help, User, and a grid icon. Logos for cirad, IRD, Bioversity, INRA, and Montpellier SupAgro are displayed. The main content area features the SouthGreen logo and a welcome message: "Welcome to GALAXY Our pre-configured and certified workflows". A central figure displays a GWAS plot with colored points representing different chromosomes. To the left, a vertical sidebar lists various bioinformatics tools categorized under NGS analyses, SNP calling, SNP analysis, and GWAS. To the right, a vertical sidebar lists datasets from the History panel, including "Copy of 'MADS rice history' shared by baptiste.guitton@cirad.fr (active items only)" and a list of numbered items from 76 to 65. A callout box in the bottom center provides instructions for finding available tools.

Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help User

Tools

search tools

Get Data Send Data

BASIC TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

SEQUENCE ANALYSIS

Gene/Protein prediction

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

Using 0%

cirad

IRD

Bioversity

INRA

Montpellier SupAgro

# SouthGreen

bioinformatics platform

## Welcome to GALAXY

### Our pre-configured and certified workflows

GWAS

SNiPlay3 GWAS workflow: Tassel-based GWAS workflow (GLM model) including population structure and correction for structure (Dereeper et al, 2015)

Input: VCF + Phenotypic tabulated file

Access workflow

Structural variations Chrom. reconstruction Metagenomics

... 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.

History

search datasets

Copy of 'MADS rice history' shared by baptiste.guitton@cirad.fr (active items only)  
37 shown  
9.45 MB

76: Rapgreen statistics

75: Reconciled Newick

74: Output species

73: Output phyloxml

72: Newick

71: Pasted Entry

70: SouthGreen visualization on data 67, data 68, and data 58

69: Get alternative gene identifiers on data 66

68: Ideeve : Identification of evolutive events based on syntetic studies on data 66

67: Differential expression on Genfam on data 66

66: Format Fasta Locus Tag on data 65

65: Format fasta header on data 64

# Galaxy

Galaxy se place directement dans le JLFM2

The screenshot shows the SouthGreen bioinformatics platform interface. At the top, there's a navigation bar with links for Analyze Data, Workflow, Shared Data, Visualization, Admin, Help, and User. Below the navigation bar, there are logos for cirad, IRD, Bioversity, and INRA. A large orange header banner reads "SouthGreen bioinformatics platform". The main content area features a "Welcome to GALAXY" message and a section titled "Our pre-configured and certified workflows". It lists several workflows: NGS analyses, SNP calling, SNP analysis, GWAS, Structural variations, Chrom. reconstruction, and Metagenomics. A "GWAS" workflow card is highlighted, showing its description: "SNiPlay3 GWAS workflow: Tassel-based GWAS workflow (GLM model) including population structure and correction for structure (Dereeper et al, 2015)", input requirements ("Input: VCF + Phenotypic tabulated file"), and a "Access workflow" button. To the left of the main content, a sidebar lists various tools under categories like BASIC TOOLS, SEQUENCE ANALYSIS, and NGS ANALYSIS. A callout box labeled "Accès direct au cluster" points to the "GWAS" workflow card. Another callout box labeled "Centralisation et partage des démarches" points to the "Chrom. reconstruction" section. A third callout box labeled "Interface et compte utilisateur indépendants du poste client" points to the "Metagenomics" section. On the right side, there's a sidebar titled "Using 0%" with a "Story" shared by baptiste.guilton@cirad.fr (active items only). It lists several items with icons for edit and delete:

- 76: Rapgreen statistics
- 75: Reconciled Newick
- 74: Output species
- 73: Output phyloxml
- 72: Newick

Below this, another sidebar titled "Members on data 66" lists items 68 through 65:

- 68: Ideven : Identification of evolutive events based on syntetic studies on data 66
- 67: Differential expression on Genfam on data 66
- 66: Format Fasta Locus Tag on data 65
- 65: Format fasta header on data 64

Callout boxes on the right side say "Utilisable depuis n'importe quel poste connecté à internet" and "Centralisation et partage des données".

# Galaxy - Vue générale

The screenshot shows the Galaxy bioinformatics platform interface. At the top, there is a navigation bar with links for Analyze Data, Workflow, Shared Data, Visualization, Admin, Help, and User. Below the navigation bar, there are several logos: cirad, IRD, Bioversity, INRA, and Montpellier SupAgro. The main content area features the SouthGreen bioinformatics platform logo and a welcome message: "Welcome to GALAXY Our pre-configured and certified workflows". On the left, there is a sidebar titled "Tools" containing a search bar and a list of pre-configured and certified workflows. On the right, there is a "History" panel showing a list of datasets with options to copy, edit, or delete them. Red circles highlight the "Tools" sidebar and the "History" panel.

Boîte à outils

Données

# Galaxy - Vue générale

Chaque programme peut être configuré précisément

Configuration d'un programme

The screenshot shows the Galaxy web interface with a red callout box highlighting the configuration area for the PhyML tool. The PhyML tool is a ML tree builder (Galaxy Version 3.0). The configuration panel includes fields for Alignment in phylip format (with a dropdown menu showing '76: Rapgreen statistics'), Data type (Nucleic acids), Evolution model (HKY85), Transition/transversion ratio ('e'), 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), and Tree topology search operation (NNI (Nearest Neighbor Interchange)). To the right, the History panel lists various datasets and operations, such as 'Copy of 'MADS rice history' shared by baptiste.guitton@cirad.fr (active items only)', '76: Rapgreen statistics', '75: Reconciled Newick', etc.

# Galaxy - Accès aux données

Les données peuvent être chargées depuis votre ordinateur local

Pour importer depuis votre ordinateur

Pour importer depuis une autre page web, ou copier-coller

# 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 currently selected and highlighted in blue), Visualization, Admin, Help, User, and a grid icon. Below the navigation bar, a red box highlights the 'Shared Data' menu. A second red box highlights the 'First Steps' folder in the left sidebar, which contains several public datasets. A third red box highlights the 'Accès aux bibliothèques' (Access to libraries) text in a red box.

DATA LIBRARIES << 0 1 2 >> showing 5 of 5 items  include deleted + + to ↗

Libraries / GenFam

name desc ...

First Steps  
GenFam Test Sets  
genome\_data  
genome\_data\_2  
TF\_data

Test our GenFam analysis workflows with the SPS family

data type size time updated (UTC)

folder 2016-05-25 09:07 AM  
folder 2016-05-25 09:06 AM  
folder 2016-07-07 05:34 PM  
folder 2016-07-07 05:43 PM  
folder 2015-07-24 07:39 AM

Using 0% Help

Accès aux bibliothèques

Données publiques

<< 0 1 2 >> showing 5 of 5 items

# Galaxy - Historiques des analyses

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

The screenshot illustrates the Galaxy web interface for managing analysis histories. On the left, the 'History' sidebar shows a list of recent datasets and their details. A red box highlights the 'Historique courant' (current history) entry. A red circle highlights the gear icon in the top right of the sidebar, which opens a context menu. The main content area is divided into two sections: 'Saved Histories' and 'Published Histories'. The 'Saved Histories' section lists various histories with details like name, datasets, tags, and creation date. A red box highlights the 'Listes des historiques personnels, et publiés' (Lists of personal and published histories) text. The 'Published Histories' section lists histories shared with others, including their names and owners.

History

search datasets

Copy of 'MADS rice history' shared by baptiste.guitton@cirad.fr (active items only)  
37 shown  
9.45 MB

76: Rapgreen statistics

75: Reconciled Newick

74: Output species

73: Output phyloxml

72: Newick

71: Pasted Entry

70: SouthGreen visualization on data 67, data 68, and data 58

69: Get alternative gene

n of evolutive events based on syntetic studies on data 66

67: Differential expression Genfam on data 66

66: Format Fasta Locus T on data 65

South Green bioinformatics platform

Historique courant

**Saved Histories**

search history names and tags

Advanced Search

Name	Datasets	Tags	Sharing	Size on Disk	Created	Last Updated
Unnamed history	0	0 Tags	0 bytes	Sep 23, 2016	Sep 23, 2016	
Copy of 'MADS rice history' shared by baptiste.guitton@cirad.fr (active items only)	53	10	0 Tags	9.4 MB	Jun 20, 2016	Jun 20, 2016
Unnamed history	0	0 Tags	0 bytes	Jun 20, 2016	Jun 20, 2016	
bordeaux_phylo_simple	15	0 Tags	2.2 MB	May 24, 2016	May 25, 2016	
proteique	23	2	0 Tags	260.9 KB	Feb 18, 2016	Feb 19, 2016
formation_phylo_nucleique_supagro_2016	10	0 Tags	1.1 MB	Feb 17, 2016	Feb 18, 2016	
formation_phylo_proteique_supagro2016	20	0 Tags	807.0 KB	Feb 18, 2016	Feb 18, 2016	
Demo galaxy	8	0 Tags	9.3 KB	Jul 23, 2015	Feb 15, 2016	

For 0 selected histories: Rename, Delete, Delete F

**Published Histories**

search name, annotation, owner, and tags

Advanced Search

Name	Annotation	Owner
Scaffremodler : Structural Variation		droc
SNIPlay		dereeper
Genome Improvement : 1. Refinement of scaffold miss-assemblies boundaries		droc
Genome Improvement : 3. Performing scaffold fusions		droc
Genome Improvement : 4. Performing scaffold junctions		droc
Scaffhunter : Markers linkage visualization		droc
Genome Improvement : 2. Identification of scaffold fusions and junctions to perform on the newly split scaffolds		droc
vs12		fallahmane1

Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help User

Listes des historiques personnels, et publiés

# Galaxy - Historiques des analyses

**HISTORY LISTS**

**Saved Histories**

Histories Shared with Me

**HISTORY ACTIONS**

Create New

Copy History

Share or Publish

Show Structure

Extract Workflow

Delete

Delete Permanently

**DATASET ACTIONS**

Copy Datasets

Dataset Security

Resume Paused Jobs

Collapse Expanded Datasets

Unhide Hidden Datasets

Delete Hidden Datasets

Purge Deleted Datasets

**DOWNLOADS**

Export Tool Citations

Export History to File

**OTHER ACTIONS**

Import from File

## Saved Histories

Search histories, names and tags

[Advanced Search](#)

Name	Datasets	Tags	Sharing	Size on Disk	Created	Last Updated
Unnamed history	0	0 Tags	0 bytes	Sep 23, 2016	Sep 23, 2016	
Copy of 'MADS rice history' shared by baptiste.guitton@cirad.fr (active items only)	53	10	0 Tags	9.4 MB	Jun 20, 2016	Jun 20, 2016
Unnamed history	0	0 Tags	0 bytes	Jun 20, 2016	Jun 20, 2016	
bordeaux_phylo_simple	15	0 Tags	2.2 MB	May 24, 2016	May 25, 2016	
proteique	23	2	0 Tags	260.9 KB	Feb 18, 2016	Feb 19, 2016
formation_phylo_nucleique_supagro_2016	10	0 Tags	1.1 MB	Feb 17, 2016	Feb 18, 2016	
formation_phylo_proteique_supagro2016	20	0 Tags	807.0 KB	Feb 18, 2016	Feb 18, 2016	
Demo galaxy	8	0 Tags	9.3 KB	Jul 23, 2015	Feb 15, 2016	

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

Panneau de gestion

South Green  
bioinformatics platform

# Galaxy - Historique des analyses

The screenshot shows the Galaxy History interface. On the left is a sidebar with 'HISTORY LISTS' and 'Saved Histories' sections. The main area displays a history titled 'Copy of 'MADS rice history' shared by baptiste.guitton@cirad.fr (active items only)'. It lists seven datasets:

ID	Name	Actions
76	Rapgreen statistics	
75	Reconciled Newick	
74	Output species	
73	Output phyloxml	
72	Newick	
71	Pasted Entry	

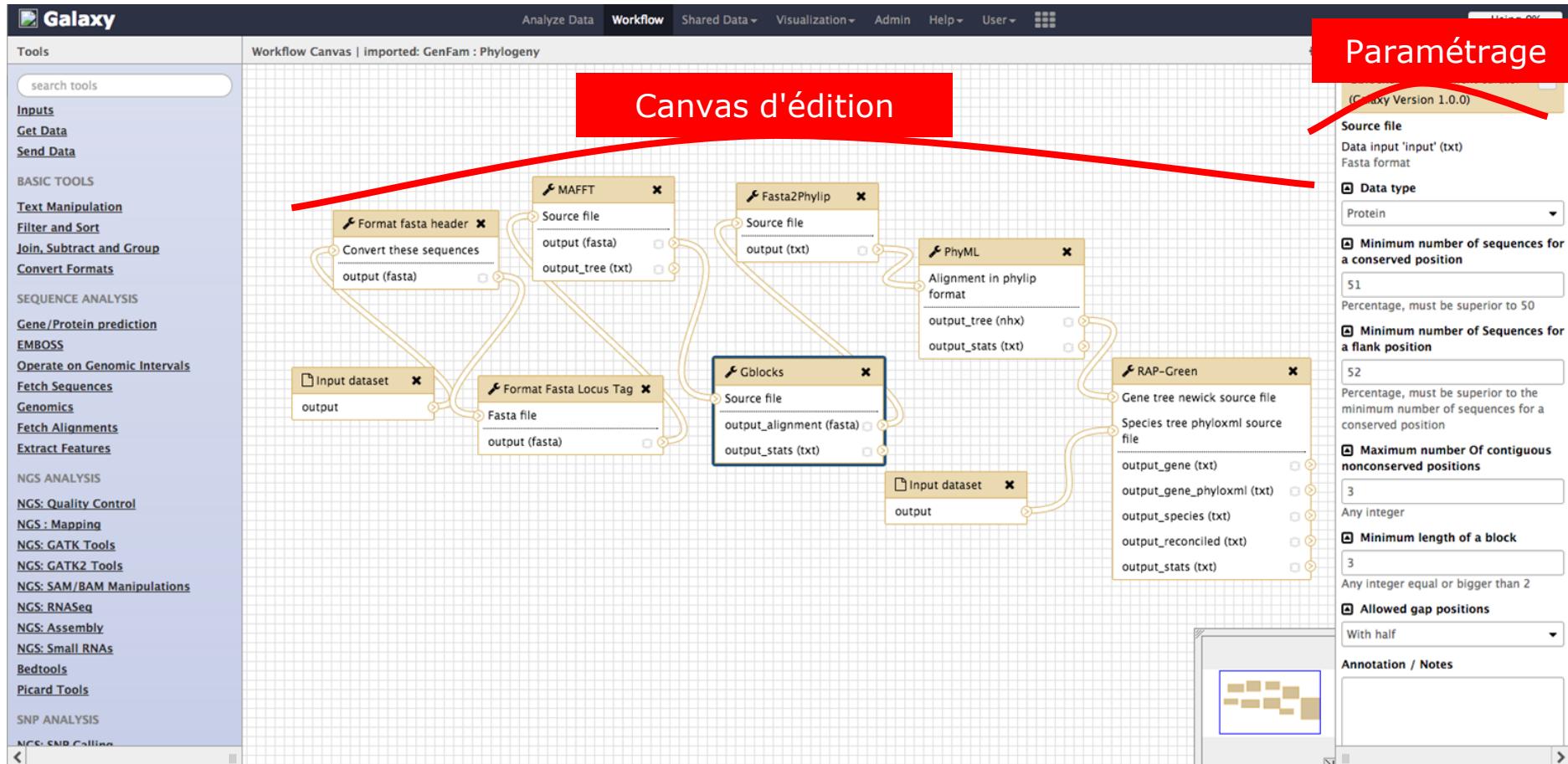
Red annotations with arrows point to specific actions:

- A red arrow points from the top right to the 'Visualiser' button, which is associated with the eye icon.
- A red arrow points from the middle right to the 'Éditer les métadonnées' button, which is associated with the pencil icon.
- A red arrow points from the bottom right to the 'Suppression temporaire' button, which is associated with the trash icon.

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

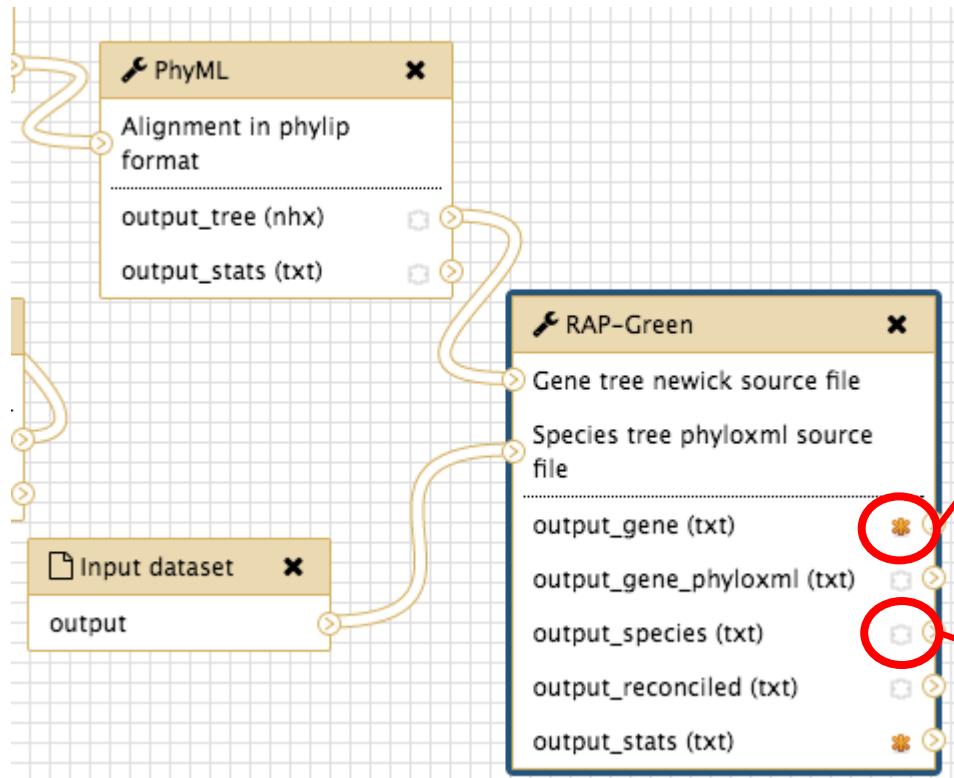
# Galaxy - Workflows

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



Le fichier apparaîtra  
dans l'historique après  
analyse

Le fichier sera masqué  
dans l'historique après  
analyse

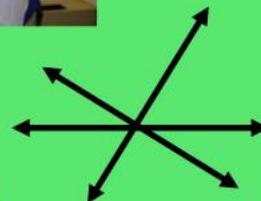
Si rien n'est coché, tout  
est coché !



---

# Galaxy au sein de Southgreen

# Plateau, plateforme, équipe de recherche



Equipe Intégration Des Données, UMR AGAP



# Galaxy, sa place dans la plateforme

## Reproductibilité des analyses

### 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 callout box points to a note about tool availability. Another callout box points to a note about new tool submissions. At the bottom, a message mentions Galaxy's support by NSF, NHGRI, and the Huck Institutes of Life Sciences.

### Utilisateur averti

### 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 'Login/Register', 'Search this site', and 'Information' (with links to Announcements, Search, Team, Publications, Seminar, Trainings, Contact, Documentation, and Links). There are also links to various databases and tools like OryGenesDB, GNPonnot, SAT, EST, SNIPlay, Haplotype, and GreenPhy 2.0.

### Utilisateur expert

A terminal window titled 'simmons@earth' shows a command-line session. The user runs 'ls' to list files in their directory, which include various C header files and shell scripts related to XTerm development.

```
[simmons@earth 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  TekPrstTbl.c  wcidwidth.c
88colres.pl     Imakefile   plink.sh   tektests
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_ap.opt
charproc.c       link_xpm.com main.c    README.os390 util.c    xterm_ap.opt
charset.c        make.com    resize.c   trace.h    xtermcf9.hin
config.guess     Makefile   resize_man UXTerm.ad  XTerm-col.ad
config.sub       Makefile.in screen.c  version.h  xterm_dat
configure        Makefile.in scrollba.c vms.c    xterm.h
configure.in     MANIFEST   menu.c   vinstall.sh xterm_io.h
cursor.c         menu.h     tabs.c    Tekparse.def xterm_log.html
data.c           menu.h
data.h           menu.h
doublechr.c     minstall.sh Tekparse.def VTparse.h
[simmons@earth xterm-XF86-4.3.0]$
```



# 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...).

## Au service des utilisateurs

Un pôle de bioanalyse, actuellement en montage.



---

# Galaxy, démonstration