





# Galaxy: ressource pour le calcul en bioinformatique

Alexis Dereeper, Jean-François Dufayard, Marilyne Summo, Gautier Sarah, Manuel Ruiz avec l'aide des formateurs précédents: Félix Homa, Stéphanie Pointet, Cédric Farcy







## La bioinformatique sans Galaxy



#### Problématique relevant de la bioinformatique









#### Problématique relevant de la bioinformatique







Je dispose d'un bioinformaticien sur place







#### Problématique relevant de la bioinformatique







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.





#### Problématique relevant de la bioinformatique









Je dispose d'une plateforme







- 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.





#### Problématique relevant de la bioinformatique









Je dispose d'une plateforme

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







- 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...



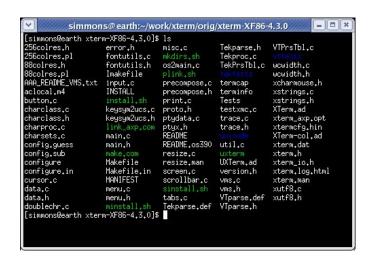
Dans cette situation, la solution la plus fréquemment employée est le

<u>JLF2M</u>: Je le fais moi-même!



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

JLF2M: Je le fais moi-même!

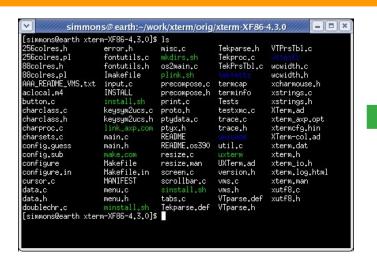


Une invite de commandes





Un cluster de calcul



Communication

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







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

#### Command line parameters

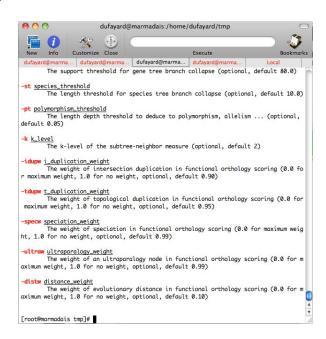
All program parameters can be entered in the command line. The first parameter must always be the name of the alignment file or the pathnames file. If this is the only given parameter, the menu of the program is activated. For example:

#### 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	

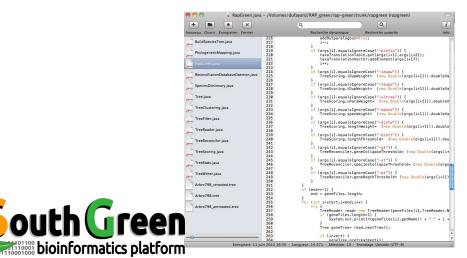


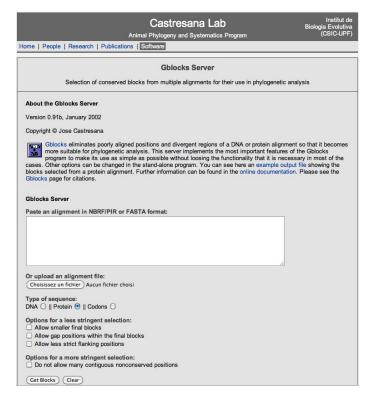




#### Construire une analyse complète

- 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.



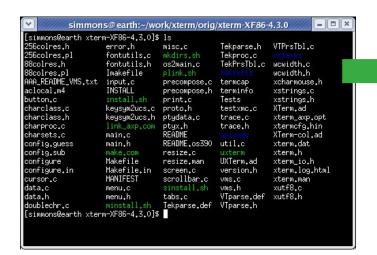




# Galaxy, rôle et utilisation?

#### Galaxy

#### Galaxy se place directement dans le JLFM2



- Galaxy

#### Communication







#### Galaxy

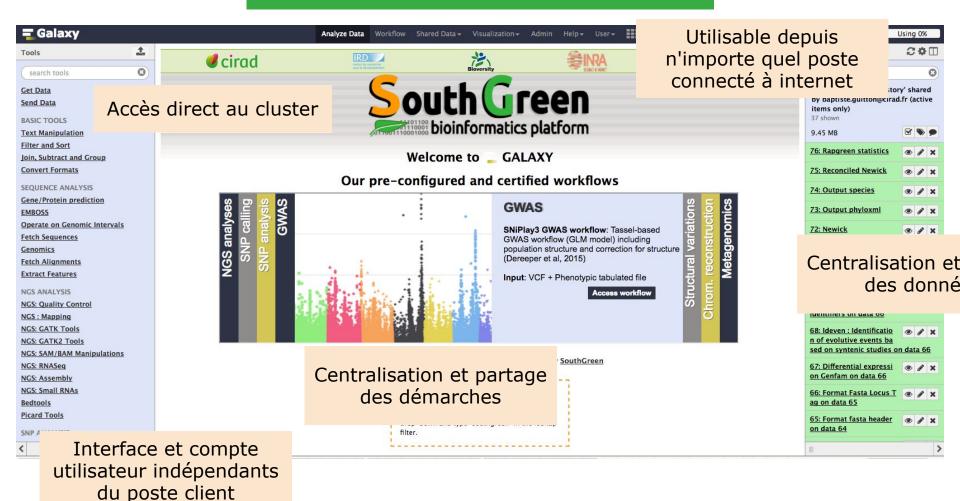
#### Galaxy se place directement dans le JLFM2





#### Galaxy

#### Galaxy se place directement dans le JLFM2





#### Galaxy - Vue générale

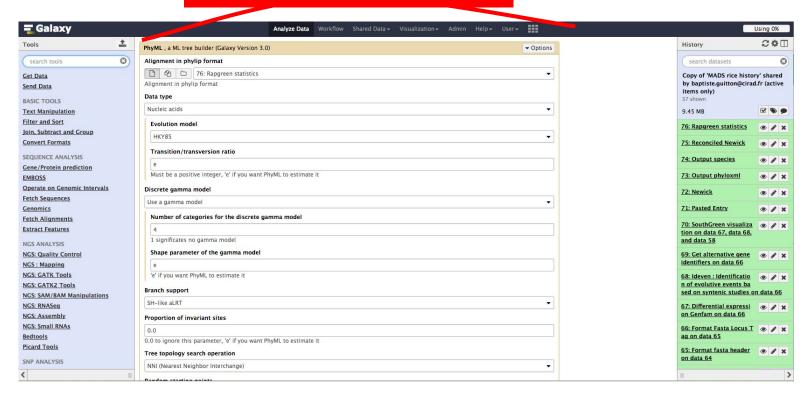




#### Galaxy - Vue générale

#### Chaque programme peut être configuré précisément

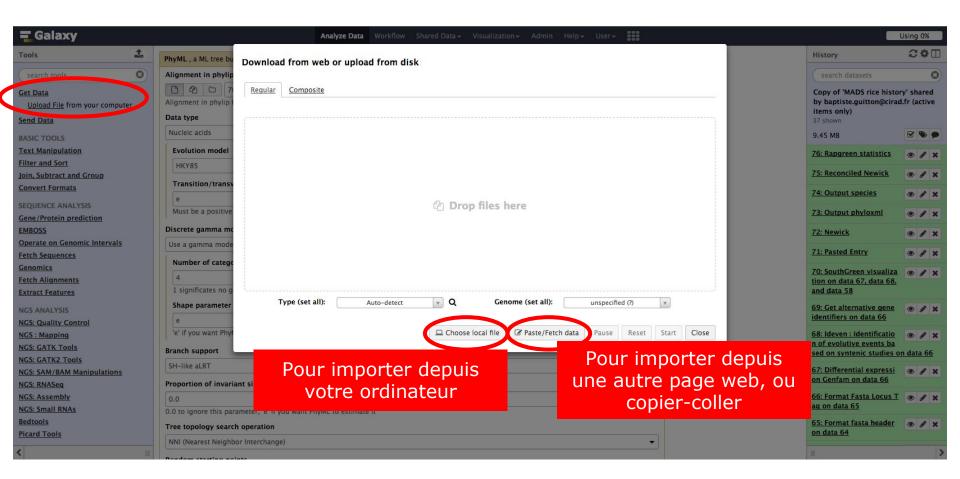
#### Configuration d'un programme





#### Galaxy - Accès aux données

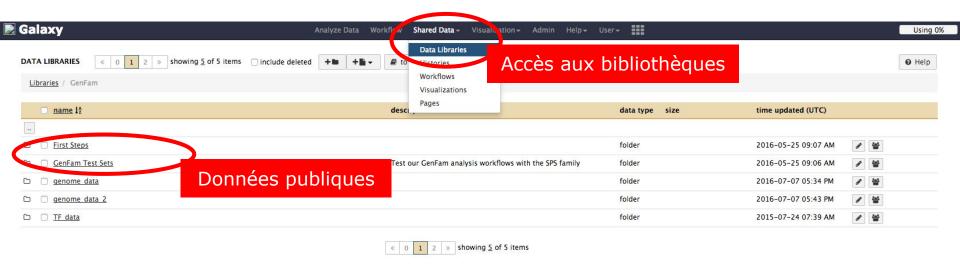
Les données peuvent être chargées depuis votre ordinateur local





#### Galaxy - accès aux données

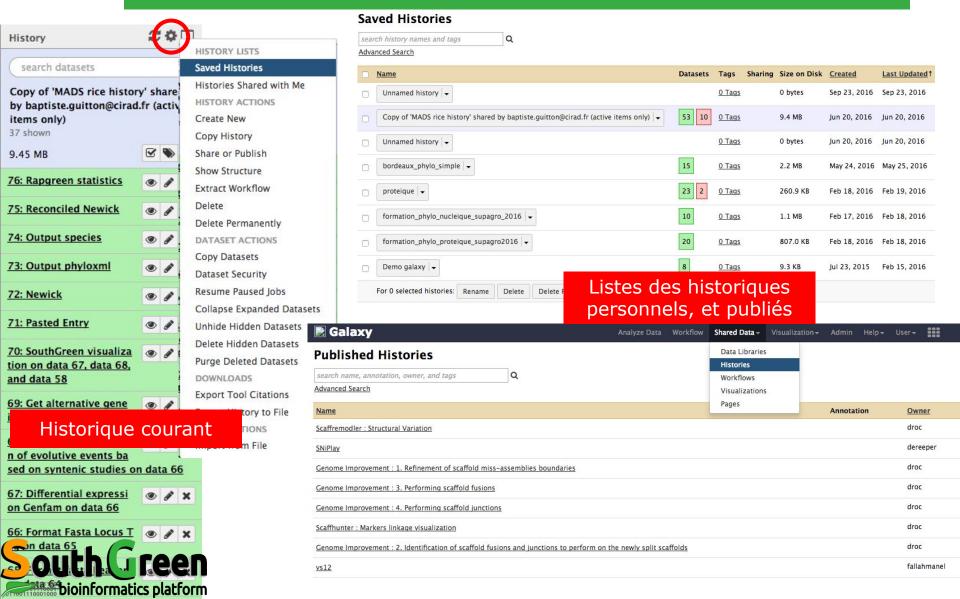
Les données peuvent être chargées depuis des bibliothèques partagées



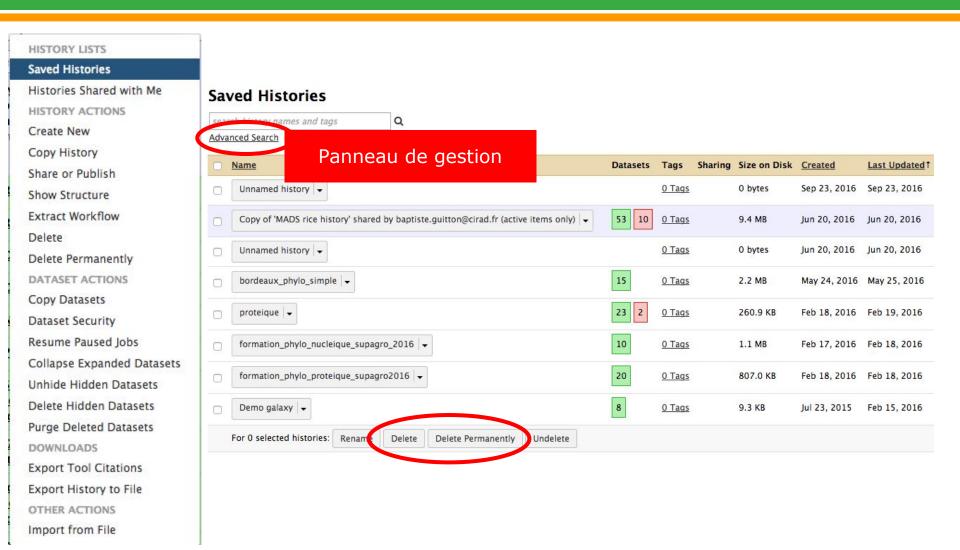


#### Galaxy - Historiques des analyses

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



#### Galaxy - Historiques des analyses





#### Galaxy - Historique des analyses

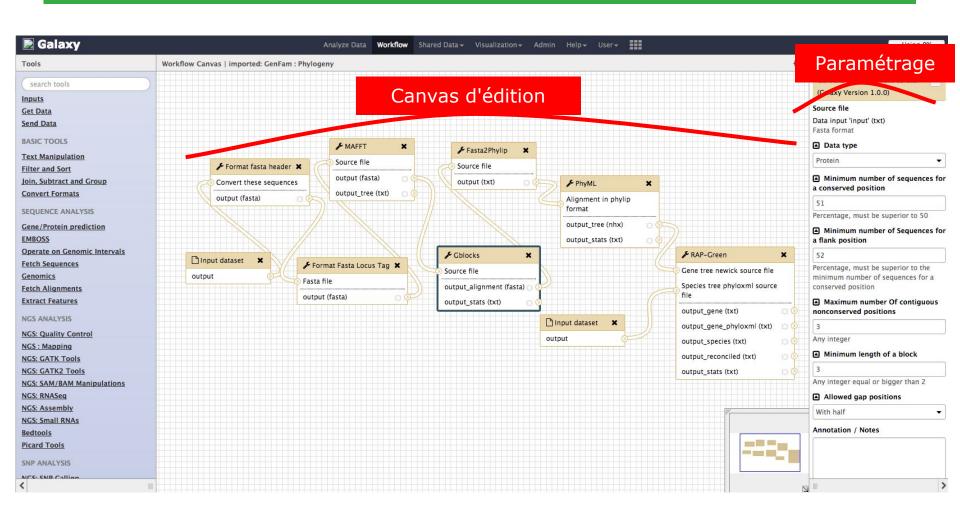


La notion d'historiques multiples est très importante pour gérer ses données et résultats



#### Galaxy - Workflows

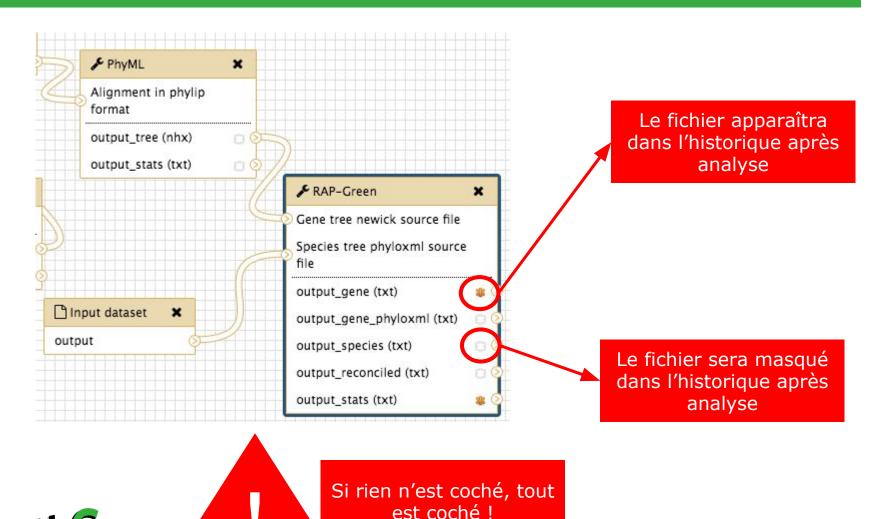
Un workflow est un enchaînement d'outils paramétrés, ils sont partageables et publiables





#### Galaxy - Workflows

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



bioinformatics platform



## Galaxy au sein de Southgreen

#### Galaxy, sa place dans la plateforme

#### Reproductibilité des analyses

#### Données à analyser **Galaxy** Analyze Data **e** cirad search tools Recently Used Get Data outh reen 20: PhyML on data 18 @ X X Send Data TOOLS Convert Formats Evolution Welcome to TGALAXY 17: Gblocks on data 15 Filter and Sort at your disposal as part of the services provided by 16: Gblocks on data 15 Gene/Protein prediction SouthGreen 15: MAFFT on data 14 @ X X NGS: Quality Control NGS: Mapping 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. 14: new.fasta NGS: SAM/BAM Manipulations

addressed to admin.bioinfo@cirad.fr

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

Données analysées et explorables



WARHAWK SERVER CLUSTER

Utilisateur averti

NGS: SNP Detection

**Protein Structures** 

Sequence comparison UNTESTED TOOLS Text Manipulation Filter and Sort

Join, Subtract and Group Convert Formats Extract Features

Fetch Sequences Fetch Alignments Operate on Genomic Intervals

simmons@earth:~/work/xterm/orig/xterm-XF86-4.3.0 [simmons@earth xterm-XF86-4.3.0]\$ 1s 256colres.h error.h misc.c Tekparse.h VTPrsTbl.c 256colres.pl fontutils.c Tekproc.c mkdirs.sh 88colres.h fontutils.h os2main.c TekPrsTbl.c wcwidth.c 88colres.pl AAA\_README\_VMS.txt Imakefile plink.sh wcwidth.h precompose.c xcharmouse.h input.c INSTALL termcap aclocal.m4 precompose.h terminfo xstrings.c button.c Tests xstrings.h print.c keysym2ucs.c charclass.c testxmc.c XTerm.ad proto.h charclass,h keysym2ucs.h ptydata.c trace.c xterm\_axp.opt charproc.c ptyx.h trace.h xtermcfg.hin README charsets.c XTerm-col.ad main.c config.guess main.h README.os390 util.c xterm.dat config.sub resize.c xterm.h Makefile configure resize.man UXTerm.ad xterm\_io.h Makefile.in xterm.log.html configure.in screen.c version.h cursor.c MANIFEST scrollbar.c VMS.C xterm.man menu.c vms.h xutf8.c VTparse.def xutf8.h menu.h tabs.c Tekparse.def [simmons@earth xterm-XF86-4.3.0]\$

Utilisateur expert





# Galaxy, démonstration