

UTILISATION DE LA PLATEFORME WEB D'ANALYSE DE DONNÉES GALAXY

Yvan Le Bras

yvan.le-bras@mnhn.fr

Marie Delannoy, Marianne Linares, Grégoire Loïs, Romain Julliard

@Yvan2935

<°))))><

Projet 65 Millions d'observateurs

Station de Biologie marine Concarneau --- MNHN EGB/CESCO UMR 7204 Paris

UTILISATION DE LA PLATEFORME WEB D'ANALYSE DE DONNÉES GALAXY

Yvan Le Bras <°)))))><
@Yvan2935

Marie Delannoy

Marianne Linares

Grégoire Loïs

Romain Julliard

CESCO UMR 7204

Projet 65 Millions d'observateurs

Station de biologie marine de Concarneau

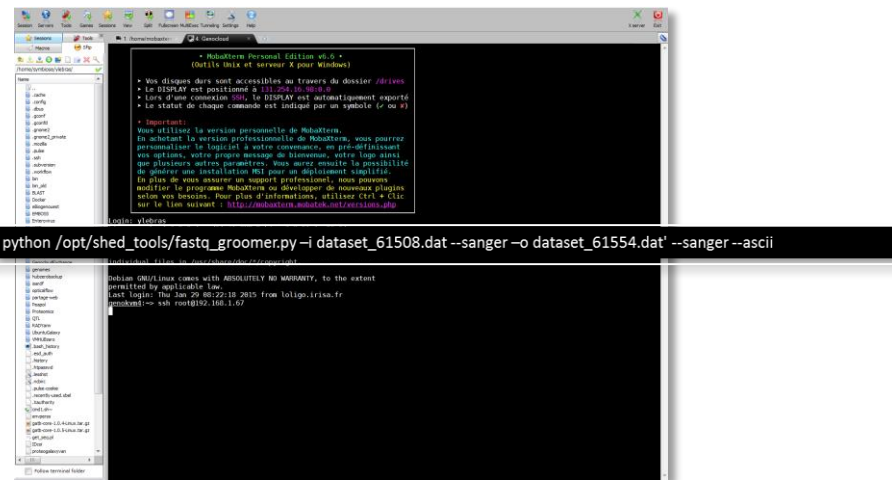


INTRODUCTION À GALAXY

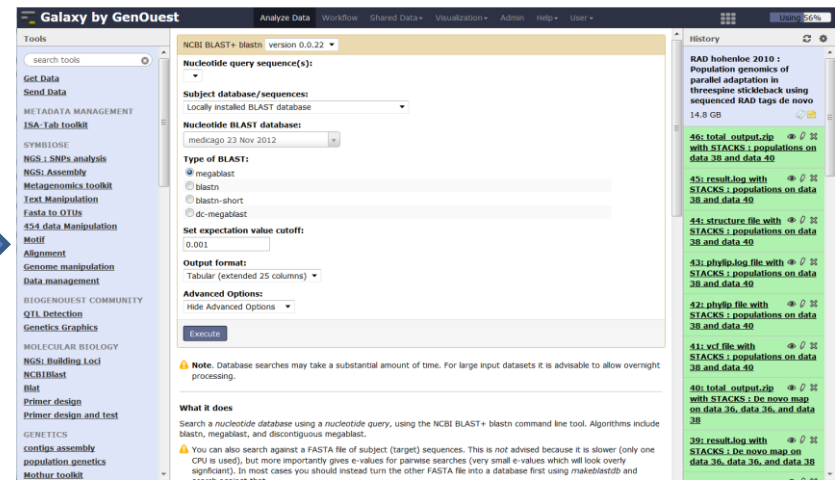
Concepts, Principe, Présentation

Concept de base

- Portail web d'analyse de données initialement génétique
- Interface web permettant d'utiliser les outils utilisable en ligne de commandes sans coder

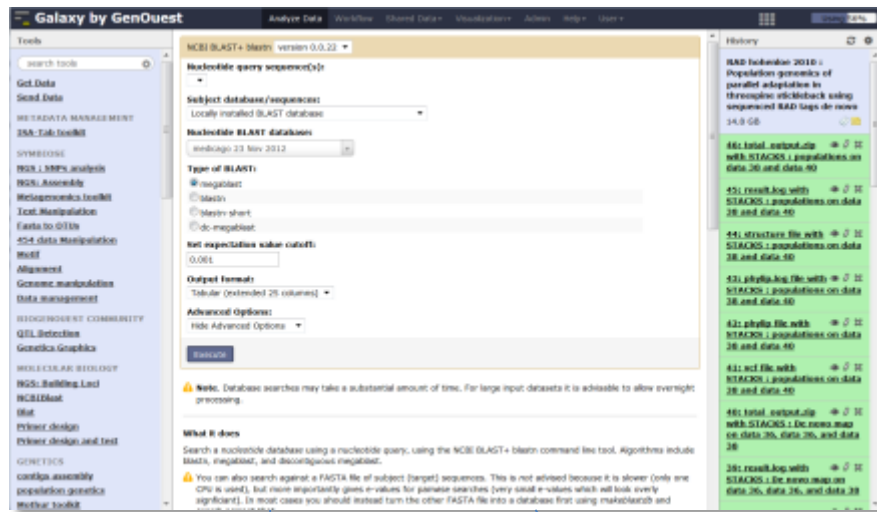


```
python /opt/shed_tools/fasto_groomer.py -i dataset_61508.dat --sanger -o dataset_61554.dat --sanger --ascii
```



- Faciliter la réanalyse
- Optimiser les ressources de traitements (calcul, stockage et réseaux)

Principe de fonctionnement



Ressources de stockage

- espace partagé
- poste de travail
- cloud ...

Ressources de calcul

- cluster de calcul
- poste de travail
- cloud ...

- 1-Téléversement/upload de la donnée
- 2-Recherche et utilisation d'un outil/script
- 3-si besoin de données externes, récupération
- 4-envoi de l'analyse sur calculateur
- 5-récupération des résultats
- 6-Enregistrement des données générées

Outils Galaxy :
logiciels éco-
informatiques,
scripts R, ...

- script R
- Tadarida
- ...

Banques de
données
externe

- habitats
- climat
- géologie ...

Présentation

Historique courant

Barre de menu

Analyse / workflow / Shared data / visualisation

Panneau d'outils

The screenshot displays the Galaxy by GenOuest web interface. On the left is a 'Tools' sidebar with a search bar and a list of tool categories including 'Get Data', 'Send Data', 'METADATA MANAGEMENT', 'SYMBIOSE', 'NGS: SNPs analysis', 'NGS: Assembly', 'Metagenomics toolkit', 'Text Manipulation', 'Fasta to OTUs', '454 data Manipulation', 'Motif', 'Alignment', 'Genome manipulation', 'Data management', 'BIOGENOUEST COMMUNITY', 'QTL Detection', 'Genetics Graphics', 'MOLECULAR BIOLOGY', 'NGS: Building Loci', 'NCBIblast', 'Blat', 'Primer design', 'Primer design and test', 'GENETICS', 'contigs assembly', 'population genetics', and 'Mothur toolkit'. The main area shows the 'NCBI BLAST+ blastn' tool configuration. The 'Nucleotide query sequence(s):' field is empty. The 'Subject database/sequences:' is set to 'Locally installed BLAST database'. The 'Nucleotide BLAST database:' is set to 'medicago 23 Nov 2012'. The 'Type of BLAST:' is set to 'megablast'. The 'Set expectation value cutoff:' is set to '0.001'. The 'Output format:' is set to 'Tabular (extended 25 columns)'. The 'Advanced Options:' section is collapsed. An 'Execute' button is visible. Below the configuration, a note states: 'Note. Database searches may take a substantial amount of time. For large input datasets it is advisable to allow overnight processing.' and a section titled 'What it does' explains the tool's function. On the right, a 'History' panel shows a list of saved jobs, including 'RAD hohenloe 2010 : Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags de novo' and several other jobs related to 'STACKS : populations on data 38 and data 40'.

Espace d'affichage des informations

Formulaire d'un Outil, donnée, historiques sauvegardés, ...

Présentation : Mode analyse

The screenshot displays the GenOuest Bioinformatics platform interface. The top navigation bar includes the 'Galaxy' logo and a menu with 'Analyze Data' (highlighted), 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. A status bar on the right indicates 'Using 52%'. The left sidebar lists various tool categories: DATA TRANSFER (Get Data, Send Data, Archive tools), SYMBIOSE (Metagenomics benchmark, Text Manipulation, Fastq to OTUs, NGS: 454 data Manipulation, Motif, NGS: Alignment, NGS: Assembly, NGS: Genome manipulation, Systems biology, Data management, GenOuest: Mothur toolkit, KLAST), COLIBREAD (Colib' read tool suite), BIOGENOUEST COMMUNITY (OTL Detection), GENOMICS (Alignment), GENETICS (Primer design, NGS: Contigs assembly, Population genetics), and a search bar. The main content area features a green banner stating 'GenOuest Galaxy server is running' and a blue information box with the email 'support@genouest.org'. The central text reads 'GenOuest Bioinformatics platform' followed by 'Development, expertise and resources for bioinformatics'. The bottom of the main area contains logos for IRISA, Inria, Biogenouest, and Région Bretagne. The right sidebar shows an 'History' panel with an 'Unnamed history' (empty) and a message: 'This history is empty. You can load your own data or get data from an external source'. A blue information box at the bottom of the main area states: 'If you use this instance to publish, please cite :'. The bottom navigation bar includes a left arrow, a search bar, and a right arrow.

Galaxy Analyze Data Workflow Shared Data Visualization Help User Using 52%

Tools

search tools

DATA TRANSFER

[Get Data](#)

[Send Data](#)

[Archive tools](#)

SYMBIOSE

[Metagenomics benchmark](#)

[Text Manipulation](#)

[Fastq to OTUs](#)

[NGS: 454 data Manipulation](#)

[Motif](#)

[NGS: Alignment](#)

[NGS: Assembly](#)

[NGS: Genome manipulation](#)

[Systems biology](#)

[Data management](#)

[GenOuest: Mothur toolkit](#)

[KLAST](#)

COLIBREAD

[Colib' read tool suite](#)

BIOGENOUEST COMMUNITY

[OTL Detection](#)

GENOMICS

[Alignment](#)

GENETICS

[Primer design](#)

[NGS: Contigs assembly](#)

[Population genetics](#)

GenOuest Galaxy server is running

For any question or tools request, please send an email at support@genouest.org

GenOuest Bioinformatics platform

Development, expertise and resources for bioinformatics

IRISA Inria Biogenouest Région BRETAGNE

History

search datasets

Unnamed history (empty)

This history is empty. You can load your own data or get data from an external source

If you use this instance to publish, please cite :

Présentation : Mode analyse

Téléverser
(uploader)
des
données

The screenshot displays the Galaxy/GenQuest web interface. A central dialog box titled "Download data directly from web or upload files from your disk" is open. The dialog contains a large dashed box for file uploads and a section for file type and genome selection. Below the dialog, a list of datasets is visible on the right side of the interface.

Dialog Box Content:

Download data directly from web or upload files from your disk

You can Drag & Drop files into this box.

Type (set all): Genome (set all):

Dataset List (Right Panel):

- Colib' read Comet : CCompare Multiple METagenomes (96.6 KB)
- 19: additional file with Decompress an archive (matrix_plain)
- 18: additional file with Decompress an archive (matrix_percentage)
- 17: additional file with Decompress an archive (matrix_normalized)
- 15: decompress_an_archive.log
- 14: comet on data 6: heatmap_plain.png
- 13: comet on data 6: heatmap_percentage.png
- 12: comet on data 6: heatmap_normalized.png
- 11: comet on data 6: matrix.zip
- 10: comet on data 6: dendrogram.png
- 9: comet on data 6: logs.zip
- 8: comet on data 6: vector.zip
- 7: comet on data 6: comet.log
- 6: read_sets : Prepare comet on data 3, data 1, and others
- 5: E.fa
- 4: D.fa
- 3: C.fa

Présentation : Mode analyse

Téléverser
(uploader)
des
données
venant de
son PC en
local (<2Go
par fichier)

Galaxy / GenOuest

Analyze Data Workflow Shared Data Visualization Admin Help User

Tools

search tools

DATA TRANSFER

- Get Data
- Send Data
- Archive tools

SYMBIOSE

- Metagenomics benchmark
- Text Manipulation
- Fasta to OTUs
- NGS: 454 data Manipulation
- Motif
- NGS: Alignment
- NGS: Assembly
- NGS: Genome manipulation
- Systems biology
- Data management
- GenOuest: Mothur toolkit
- KLAST

COLIBREAD

- Colib' read tool suite

BIOGENOUEST COMMUNITY

- OTL Detection

GENOMICS

- Alignment

GENETICS

- Primer design
- NGS: Contigs assembly
- Population genetics

PHYLOGENY

- Text Manipulation
- Next Generation Quality
- Phylogenetic Tree

Alignment

- Blat
- Phylostatistics

METADATA MANAGEMENT

- ISA-Tab toolkit

GALAXY TOOLS

- Lift-Over

Download data directly from web or upload files from your disk

You added 5 file(s) to the queue. Add more files or click 'Start' to proceed.

Name	Size	Type	Genome	Settings	Status
Galaxy5-[E.fa].fasta	1.2 MB	Auto-det...	Q	unspecified (?)	
Galaxy4-[D.fa].fasta	1.4 MB	Auto-det...	Q	unspecified (?)	
Galaxy3-[C.fa].fasta	1.2 MB	Auto-det...	Q	unspecified (?)	
Galaxy2-[B.fa].fasta	1.4 MB	Auto-det...	Q	unspecified (?)	
Galaxy1-[A.fa].fasta	1.4 MB	Auto-det...	Q	unspecified (?)	

Type (set all): Auto-detect Q Genome (set all): unspecified (?)

Choose local file Choose FTP file Paste/Fetch data Start Pause Reset Close

History

search datasets

Commet metagenomics datasets comparison

0 bytes

This history is empty. You can load your own data or get data from an external source

Présentation : Mode analyse

Téléverser
(uploader)
des
données
envoyées
précédem
ent sur le
serveur FTP
de Galaxy
(pour les
fichiers
>2Go)

Galaxy / GenOuest

Analyze Data Workflow Shared Data Visualization Admin Help User

Tools

search tools

DATA TRANSFER

- Get Data
- Send Data
- Archive tools

SYMBIOSE

- Metagenomics benchmark
- Text Manipulation
- Fasta to OTUs
- NGS: 454 data Manipulation
- Motif
- NGS: Alignment
- NGS: Assembly
- NGS: Genome manipulation
- Systems biology
- Data management
- GenOuest: Mothur toolkit
- KLAST

COLIBREAD

- Colib' read tool suite

BIOPENQUEST COMMUNITY

- OTL Detection

GENOMICS

- Alignment

GENETICS

- Primer design
- NGS: Contigs assembly
- Population genetics

PHYLOGENY

- Text Manipulation
- Next Generation Quality
- Phylogenetic Tree

Alignment

- Blat
- Phylostatistics

METADATA MANAGEMENT

- ISA-Tab toolkit

GALAXY TOOLS

- Lift-Over

Download data directly from web or upload files from your disk

You added 5 file(s) to the queue. Add more files or click 'Start' to proceed.

Name	Size	Type	Genome	Settings	Status
Galaxy5-[E.fa].fasta	1.2 MB	Auto-det...	Q	unspecified (?)	
Galaxy4-[D.fa].fasta	1.4 MB	Auto-det...	Q	unspecified (?)	
Galaxy3-[C.fa].fasta	1.2 MB	Auto-det...	Q	unspecified (?)	
Galaxy2-[B.fa].fasta	1.4 MB	Auto-det...	Q	unspecified (?)	
Galaxy1-[A.fa].fasta	1.4 MB	Auto-det...	Q	unspecified (?)	

Type (set all): Auto-detect Q Genome (set all): unspecified (?)

Choose local file Choose FTP file Paste/Fetch data Start Pause Reset Close

History

search datasets

Commet metagenomics datasets comparison

0 bytes

This history is empty. You can load your own data or get data from an external source

Présentation : Mode analyse

Téléverser
(uploader)
des
données en
les rentrant
directement
à la main ou
par
copie/coller
(petits jeux
de données
textes) OU
en copiant
l'URL de la
page web
pointant sur
les données

Galaxy / GenOuest

Analyze Data Workflow Shared Data Visualization Admin Help User

Tools

search tools

DATA TRANSFER

Get Data

Send Data

Archive tools

SYMBIOSE

Metagenomics benchmark

Text Manipulation

Fasta to OTUs

NGS: 454 data Manipulation

Motif

NGS: Alignment

NGS: Assembly

NGS: Genome manipulation

Systems biology

Data management

GenOuest: Mothur toolkit

KLAST

COLIBREAD

Colib' read tool suite

BIOGENOUEST COMMUNITY

OTL Detection

GENOMICS

Alignment

GENETICS

Primer design

NGS: Contigs assembly

Population genetics

PHYLOGENY

Text Manipulation

Next Generation Quality

Phylogenetic Tree

Alignment

Blat

Phylostatistics

METADATA MANAGEMENT

ISA-Tab toolkit

GALAXY TOOLS

Lift-Over

Download data directly from web or upload files from your disk

You added 5 file(s) to the queue. Add more files or click 'Start' to proceed.

Name	Size	Type	Genome	Settings	Status
Galaxy5-[E.fa].fasta	1.2 MB	Auto-det...	Q	unspecified (?)	
Galaxy4-[D.fa].fasta	1.4 MB	Auto-det...	Q	unspecified (?)	
Galaxy3-[C.fa].fasta	1.2 MB	Auto-det...	Q	unspecified (?)	
Galaxy2-[B.fa].fasta	1.4 MB	Auto-det...	Q	unspecified (?)	
Galaxy1-[A.fa].fasta	1.4 MB	Auto-det...	Q	unspecified (?)	

Type (set all): Auto-detect Q Genome (set all): unspecified (?)

Choose local file Choose FTP file Paste/Fetch data Start Pause Reset Close

History

search datasets

Commet metagenomics datasets comparison

0 bytes

This history is empty. You can load your own data or get data from an external source

Présentation : Mode analyse

Ici nous téléversons des données venant de notre PC en local (<2Go par fichier)

The screenshot displays the Galaxy/GenOuest web interface. A modal window titled "Download data directly from web or upload files from your disk" is open. It shows a table of 5 files added to the queue, all of type "Auto-det..." and size 1.2 MB or 1.4 MB. The modal includes a "Choose local file" button, which is highlighted with a red box and a red arrow pointing to it from the text on the left. Other buttons include "Choose FTP file", "Paste/Fetch data", "Start", "Pause", "Reset", and "Close". The background shows the Galaxy/GenOuest navigation menu on the left and a "History" panel on the right.

Download data directly from web or upload files from your disk

You added 5 file(s) to the queue. Add more files or click 'Start' to proceed.

Name	Size	Type	Genome	Settings	Status
Galaxy5-[E.fa].fasta	1.2 MB	Auto-det...	unspecified (?)	⚙️	🗑️
Galaxy4-[D.fa].fasta	1.4 MB	Auto-det...	unspecified (?)	⚙️	🗑️
Galaxy3-[C.fa].fasta	1.2 MB	Auto-det...	unspecified (?)	⚙️	🗑️
Galaxy2-[B.fa].fasta	1.4 MB	Auto-det...	unspecified (?)	⚙️	🗑️
Galaxy1-[A.fa].fasta	1.4 MB	Auto-det...	unspecified (?)	⚙️	🗑️

Type (set all): Auto-detect Q Genome (set all): unspecified (?)

Choose local file Choose FTP file Paste/Fetch data Start Pause Reset Close

Présentation : Mode analyse

Téléverser
(uploader)
des
données

The screenshot displays the Galaxy/GenOuest web interface. A modal dialog box is open, titled "Download data directly from web or upload files from your disk". The dialog box contains a table of files to be uploaded, with columns for Name, Size, Type, Genome, Settings, and Status. The files listed are Galaxy5-[E.fa].fasta, Galaxy4-[D.fa].fasta, Galaxy3-[C.fa].fasta, Galaxy2-[B.fa].fasta, and Galaxy1-[A.fa].fasta. The 'Start' button is highlighted with a red box.

Name	Size	Type	Genome	Settings	Status
Galaxy5-[E.fa].fasta	1.2 MB	Auto-det...	unspecified (?)		100%
Galaxy4-[D.fa].fasta	1.4 MB	Auto-det...	unspecified (?)		100%
Galaxy3-[C.fa].fasta	1.2 MB	Auto-det...	unspecified (?)		100%
Galaxy2-[B.fa].fasta	1.4 MB	Auto-det...	unspecified (?)		100%
Galaxy1-[A.fa].fasta	1.4 MB	Auto-det...	unspecified (?)		100%

Below the table, there are dropdown menus for "Type (set all):" and "Genome (set all):". At the bottom of the dialog box, there are buttons for "Choose local file", "Choose FTP file", "Paste/Fetch data", "Start", "Pause", "Reset", and "Close". The "Start" button is highlighted with a red box.

Présentation : Mode analyse

1er jeu de données de
l'historique courant nommé
« commet metagenomics ... »

The screenshot displays the Galaxy/GenOuest web interface in the 'Analyze Data' mode. The top navigation bar includes links for 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. The left sidebar contains a 'Tools' section with a search bar and various tool categories: DATA TRANSFER, SYMBIOSE, NGS: 454 data Manipulation, NGS: Alignment, NGS: Assembly, NGS: Genome manipulation, Systems biology, Data management, GenOuest: Mothur toolkit, KLASI, COLIBREAD, BIOGENOUEST COMMUNITY, QTL Detection, GENOMICS, GENETICS, PHYLOGENY, METADATA MANAGEMENT, and GALAXY TOOLS. The central workspace features a warning message: 'This dataset is large and only the first megabyte is shown below.' Below this, a FASTA file named '1: Galaxy5-[E.fa].fasta' is displayed, showing sequence data. The right sidebar contains a 'History' panel with a search bar and a list of datasets. The selected dataset is '1: Galaxy5-[E.fa].fasta', which is 10,000 sequences in FASTA format. A 'View data' button is visible next to the dataset name.

Présentation : Mode analyse

Visualisation du 1er jeu de données

The screenshot displays the Galaxy/GenOuest web interface. The top navigation bar includes links for 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. The left sidebar lists various tool categories such as 'DATA TRANSFER', 'SYMBIOSE', 'NGS: Alignment', 'NGS: Assembly', 'NGS: Genome manipulation', 'Systems biology', 'Data management', 'GenOuest: Mothur toolkit', 'KLAIST', 'COLIBREAD', 'BIOGENOUEST COMMUNITY', 'QTL Detection', 'GENOMICS', 'GENETICS', 'PHYLOGENY', 'METADATA MANAGEMENT', and 'GALAXY TOOLS'. The central panel shows a dataset with a warning message: 'This dataset is large and only the first megabyte is shown below.' Below this, the sequence data is displayed in a text format. A blue box highlights the sequence content, labeled 'Contenu du fichier'. The right sidebar shows a 'History' section with a list of datasets, including 'Galaxy1-[A.fal.fasta]', 'Galaxy2-[B.fal.fasta]', 'Galaxy3-[C.fal.fasta]', 'Galaxy4-[D.fal.fasta]', and 'Galaxy5-[E.fal.fasta]'. A blue box highlights this list, labeled 'Prévisualisation'.

Galaxy / GenOuest

Analyze Data Workflow Shared Data Visualization Admin Help User

Tools

search tools

DATA TRANSFER

Get Data

Send Data

Archive tools

SYMBIOSE

Metagenomics benchmark

Text Manipulation

Fasta to OTUs

NGS: 454 data Manipulation

Motif

NGS: Alignment

NGS: Assembly

NGS: Genome manipulation

Systems biology

Data management

GenOuest: Mothur toolkit

KLAIST

COLIBREAD

Colib'read tool suite

BIOGENOUEST COMMUNITY

QTL Detection

GENOMICS

Alignment

GENETICS

Primer design

NGS: Contigs assembly

Population genetics

PHYLOGENY

Text Manipulation

Next Generation Quality

Phylogenetic Tree

Alignment

Blat

Phylostatistics

METADATA MANAGEMENT

ISA-Tab toolkit

GALAXY TOOLS

Lift-Over

This dataset is large and only the first megabyte is shown below.

Show all Save

>0
AAAAAAAAACGGAAGTCTCTTGCAGTATGCACAGGAGATCGGCGGAACCGGAATCCCCGTTGCGACTGTGACTCTTGGAGCGTGCCTAATCGGTCGGTAACTTTAATT
>1
AAAAAGCGGAATCACTGAAGCGTCAITTCCTCTCGAGAAAAGTTTACCAITGTTCTAATCTTGACCCCTACGAATCAAGGTTCCGTTAGCGTGTCTTATAGAACAAT
>2
AAAAATACAGTATATGCCCGAAGTACGCAATCCGATCCCAATAATCAATCAACCGTGTGATGGTCAGATATCTTCCGCGGATTTGGGGGACGCTACTTCGTGG
>3
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>4
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>5
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>6
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>7
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>8
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>9
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>10
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>11
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>12
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>13
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>14
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>15
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>16
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>17
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>18
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>19
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>20
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>21
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>22
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>23
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>24
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>25
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>26
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>27
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>28
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>29
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>30
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>31
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>32
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>33
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>34
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>35
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>36
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>37
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>38
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>39
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>40
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>41
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>42
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>43
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>44
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>45
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>46
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>47
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>48
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>49
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA
>50
AAAAACGAGGCTGAGCAITCTGCGACATATAGTATCCGATCAATTAATGTTGATTTAACTAAATTTGCTCACTAGCGACCCCTAAACCCGCTTACTCGAGGGCA

History

search datasets

Commet metagenomics datasets comparison

5 shown

6.3 MB

5: Galaxy1-[A.fal.fasta]

4: Galaxy2-[B.fal.fasta]

3: Galaxy3-[C.fal.fasta]

2: Galaxy4-[D.fal.fasta]

1: Galaxy5-[E.fal.fasta]

10,000 sequences

format: fasta, database: 2

uploaded fasta file

View data

Prévisualisation

Présentation : Mode analyse

Moteur de recherche dans panneau d'outils

Formulaire de l'outil

Résultat de la requête

Outil COMMET

Jeu de données d'entrée

Paramètres de l'outil

Galaxy / GenOuest

Tools

commet

Colib' read tool suite

commet Compare Multiple METagenomes

Workflows

All workflows

Read sets

1: Read sets

Dataset

5: Galaxy1-[A.fa].fasta
4: Galaxy2-[B.fa].fasta
3: Galaxy3-[C.fa].fasta
2: Galaxy4-[D.fa].fasta
1: Galaxy5-[E.fa].fasta

Accept fasta/fastq/fastq.gz

Set name

set_name

Please don't use spaces or special characters

+ Insert Read sets

Size of kmers

33

Set the length of used kmers.

Mini shared kmers

2

Minimal number of shared kmers.

Read mini length

0

Minimal length a read should have to be kept.

Mini Shannon index

0

Minimal Shannon index a read should have to be kept. Float in [0,2.32]

Reads filter options

default

by default, all reads are kept with no Ns limitation

Execute

Description

COMMET (COMpare Multiple METagenomes) provides a global similarity overview between all datasets of a large metagenomic project. Directly from non-assembled reads, all against all comparisons are performed through an efficient indexing strategy. Then, results are stored as bit vectors, a compressed representation of read files, that can be used to further combine read subsets by common logical operations. Finally, COMMET computes a clusterization of metagenomic datasets, which is visualized by dendrogram and heatmaps.

Web site

<http://colibread.inria.fr/commet/>

History

search datasets

Commet metagenomics datasets comparison

5 shown

6.3 MB

5: Galaxy1-[A.fa].fasta
4: Galaxy2-[B.fa].fasta
3: Galaxy3-[C.fa].fasta
2: Galaxy4-[D.fa].fasta
1: Galaxy5-[E.fa].fasta

10,000 sequences
format: fasta, database: 2

uploaded fasta file

```
>0
AAAAAAAAACCGAAGTCTCTTTCAGTATGACAGGAGATGTCGCGAGAA
>2
AAAAAAAAACCGAAGTCTCTTTCAGTATGACAGGAGATGTCGCGAGAA
>3
AAAAAAAAACCGAAGTCTCTTTCAGTATGACAGGAGATGTCGCGAGAA
```


Présentation : Mode analyse

Galaxy / GenOuest | Analyze Data | Workflow | Shared Data | Visualization | Admin | Help | User | Using 45%

Tools

commet

Colib' read tool suite

commet Compare Multiple METagenomes

Workflows

- All workflows

commet Compare Multiple METagenomes (Galaxy Tool Version 24.7.14) Options

Read sets

1: Read sets

Dataset

5: Galaxy1-[A.fa].fasta
4: Galaxy2-[B.fa].fasta
3: Galaxy3-[C.fa].fasta
2: Galaxy4-[D.fa].fasta
1: Galaxy5-[E.fa].fasta

Accept fasta/fastq/fastq.gz

Set name

set_name

Please don't use spaces or special characters

+ Insert Read sets

Size of kmers

33

Set the length of used kmers.

Mini shared kmers

2

Minimal number of shared kmers.

Read mini length

0

Minimal length a read should have to be kept.

Mini Shannon index

0

Minimal Shannon index a read should have to be kept. Float in [0,2.32]

Reads filter options

default

by default, all reads are kept with no Ns limitation

Execute

Description

COMMET (COMpare Multiple METagenomes™) provides a global similarity overview between all datasets of a large metagenomic project. Directly from non-assembled reads, all against all comparisons are performed through an efficient indexing strategy. Then, results are stored as bit vectors, a compressed representation of read files, that can be used to further combine read subsets by common logical operations. Finally, COMMET computes a clusterization of metagenomic datasets, which is visualized by dendrogram and heatmaps.

Web site

<http://colibread.inria.fr/commet/>

History

search datasets

Commet metagenomics datasets comparison

5 shown

6.3 MB

5: Galaxy1-[A.fa].fasta
4: Galaxy2-[B.fa].fasta
3: Galaxy3-[C.fa].fasta
2: Galaxy4-[D.fa].fasta
1: Galaxy5-[E.fa].fasta

10,000 sequences
format: fasta, database: 2

uploaded fasta file

>0
AAAAAAAAACCGAAGTCTCTTTCAGTATGCACAGGAGATCCCGAGAA
>2
AAAAAAAAACCGAAGTCTCTTTCAGTATGCACAGGAGATCCCGAGAA
>3
AAAAAAAAACCGAAGTCTCTTTCAGTATGCACAGGAGATCCCGAGAA

Paramètres par défaut, fixés, ou modification possible

Paramètres de l'outil

Présentation : Mode analyse

Galaxy / GenOuest | Analyze Data | Workflow | Shared Data | Visualization | Admin | Help | User | Using 45%

Tools

- commet
- Colib' read tool suite
 - commet Compare Multiple METagenomes
- Workflows
 - All workflows

commet Compare Multiple METagenomes (Galaxy Tool Version 24.7.14)

Read sets

1: Read sets

Dataset

- 5: Galaxy1-[A.fa].fasta
- 4: Galaxy2-[B.fa].fasta
- 3: Galaxy3-[C.fa].fasta
- 2: Galaxy4-[D.fa].fasta
- 1: Galaxy5-[E.fa].fasta

Accept fasta/fastq/fastq.gz

Set name

set_name

Please don't use spaces or special characters

+ Insert Read sets

Size of kmers

33

Set the length of used kmers.

Mini shared kmers

2

Minimal number of shared kmers.

Read mini length

0

Minimal length a read should have to be kept.

Mini Shannon index

0

Minimal Shannon index a read should have to be kept. Float in [0,2.32]

Reads filter options

default

by default, all reads are kept with no Ns limitation

Execute

Description

COMMET (COMpare Multiple METagenomes) provides a global similarity overview between all datasets of a large metagenomic project. Directly from non-assembled reads, all against all comparisons are performed through an efficient indexing strategy. Then, results are stored as bit vectors, a compressed representation of read files, that can be used to further combine read subsets by common logical operations. Finally, COMMET computes a clusterization of metagenomic datasets, which is visualized by dendrogram and heatmaps.

Web site

<http://colibread.inria.fr/commet/>

History

search datasets

Commet metagenomics datasets comparison

5 shown

6.3 MB

- 5: Galaxy1-[A.fa].fasta
- 4: Galaxy2-[B.fa].fasta
- 3: Galaxy3-[C.fa].fasta
- 2: Galaxy4-[D.fa].fasta
- 1: Galaxy5-[E.fa].fasta

10,000 sequences
format: fasta, database: 2

uploaded fasta file

```
>0
AAAAAAAAACCGAGTCTCTTTGAGTATGACAGGAGATCCCGAGAA
>2
AAAAAAAAACCGAGTCTCTTTGAGTATGACAGGAGATCCCGAGAA
>3
AAAAAAAAACCGAGTCTCTTTGAGTATGACAGGAGATCCCGAGAA
```

Présentation : Mode analyse

The screenshot displays the Galaxy / GenOuest web interface. The top navigation bar includes links for 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. The left sidebar contains a 'Tools' section with a search bar and a 'Workflows' section with a link to 'All workflows'. The main content area shows a green notification box with a checkmark icon, stating: '1 job has been successfully added to the queue - resulting in the following datasets:'. Below this, a list of datasets is displayed, each with a job ID and a description: 6: comet on data 3, data 2, and others: vector.zip; 7: comet on data 3, data 2, and others: logs.zip; 8: comet on data 3, data 2, and others: dendrogram.png; 9: comet on data 3, data 2, and others: matrix.zip; 10: comet on data 3, data 2, and others: heatmap_normalized.png; 11: comet on data 3, data 2, and others: heatmap_percentage.png; 12: comet on data 3, data 2, and others: heatmap_plain.png. A blue callout box points to the notification, containing the text 'Nouveaux jeux de données en création'. Another blue callout box points to the dataset list, containing the text 'Le « job » a été lancé sur le « calculateur » (pc perso / cluster de calcul / autre)'. The right sidebar shows a 'History' section with a search bar and a list of datasets, including 'Commet metagenomics datasets comparison' and '12 shown'.

Galaxy / GenOuest

Analyze Data Workflow Shared Data Visualization Admin Help User

Tools

commet

Colibri read tool suite

commet Compare Multiple METagenomes

Workflows

All workflows

1 job has been successfully added to the queue - resulting in the following datasets:

6: comet on data 3, data 2, and others: vector.zip

7: comet on data 3, data 2, and others: logs.zip

8: comet on data 3, data 2, and others: dendrogram.png

9: comet on data 3, data 2, and others: matrix.zip

10: comet on data 3, data 2, and others: heatmap_normalized.png

11: comet on data 3, data 2, and others: heatmap_percentage.png

12: comet on data 3, data 2, and others: heatmap_plain.png

You can check the status of queued jobs and view the resulting data from 'running' to 'finished' if completed successfully or 'error' if not.

Nouveaux jeux de données en création

Le « job » a été lancé sur le « calculateur » (pc perso / cluster de calcul / autre)

History

search datasets

Commet metagenomics datasets comparison

12 shown

6.3 MB

12: comet on data 3, data 2, and others: heatmap_plain.png

11: comet on data 3, data 2, and others: heatmap_percentage.png

10: comet on data 3, data 2, and others: heatmap_normalized.png

9: comet on data 3, data 2, and others: matrix.zip

8: comet on data 3, data 2, and others: dendrogram.png

7: comet on data 3, data 2, and others: logs.zip

6: comet on data 3, data 2, and others: vector.zip

5: Galaxy1-[A.fasta]

4: Galaxy2-[B.fasta]

3: Galaxy3-[C.fasta]

2: Galaxy4-[D.fasta]

1: Galaxy5-[E.fasta]

Présentation : Mode analyse

Galaxy / GenOuest

Tools:

Colib' read tool suite

[commet](#) Compare Multiple METagenomes

Workflows

- [All workflows](#)

Execution Log:

- 1 job has been successfully added
- 6: commet on data 3, data 2, and others: vector.zip
- 7: commet on data 3, data 2, and others: logs.zip
- 8: commet on data 3, data 2, and others: dendrogram.png
- 9: commet on data 3, data 2, and others: matrix.zip
- 10: commet on data 3, data 2, and others: heatmap_normalized.png
- 11: commet on data 3, data 2, and others: heatmap_percentage.png
- 12: commet on data 3, data 2, and others: heatmap_plain.png

You can check the status of queued jobs and view the resulting data. When a job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if it fails.

History

search datasets

Commet metagenomics datasets comparison

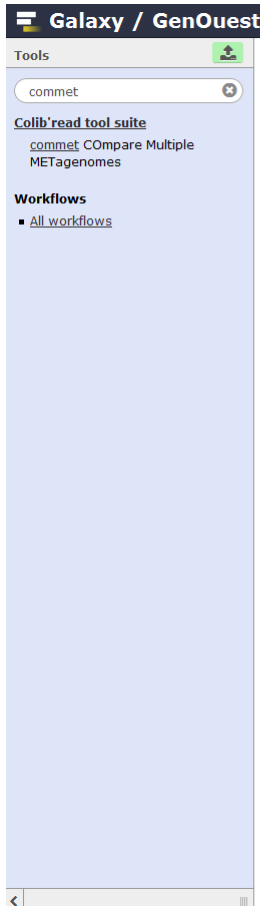
12 shown

6.3 MB

- 12: commet on data 3, data 2, and others: heatmap_plain.png
- 11: commet on data 3, data 2, and others: heatmap_percentage.png
- 10: commet on data 3, data 2, and others: heatmap_normalized.png
- 9: commet on data 3, data 2, and others: matrix.zip
- 8: commet on data 3, data 2, and others: dendrogram.png
- 7: commet on data 3, data 2, and others: logs.zip
- 6: commet on data 3, data 2, and others: vector.zip
- 5: Galaxy1-[A.fa].fasta
- 4: Galaxy2-[B.fa].fasta
- 3: Galaxy3-[C.fa].fasta
- 2: Galaxy4-[D.fa].fasta
- 1: Galaxy5-[E.fa].fasta

Présentation : Mode analyse

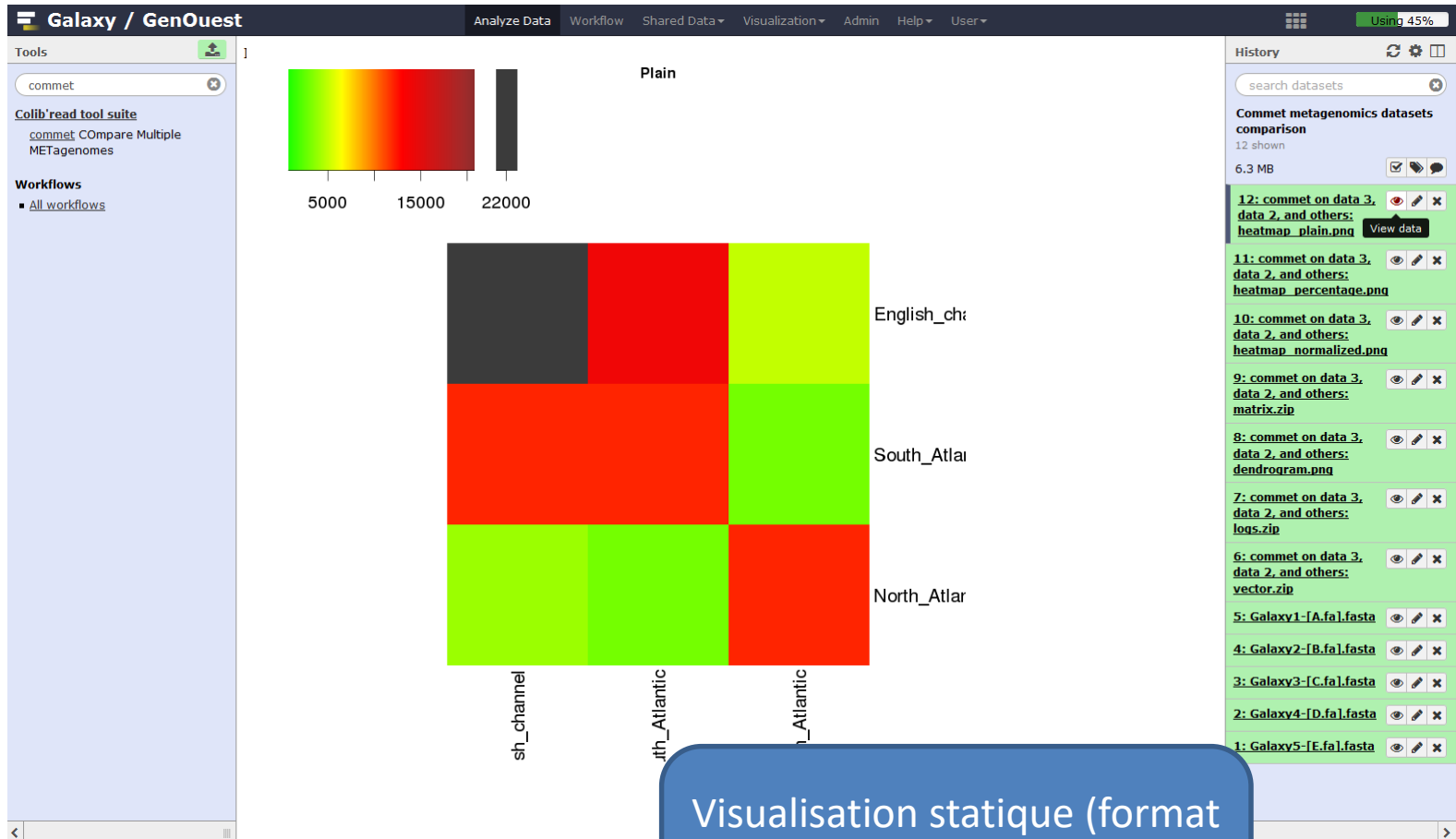
Remplissage de l'historique avec les nouveaux jeux de données générés



Le « job » est terminé et tout s'est bien passé (nouveaux jeux de données vert sinon rouge)



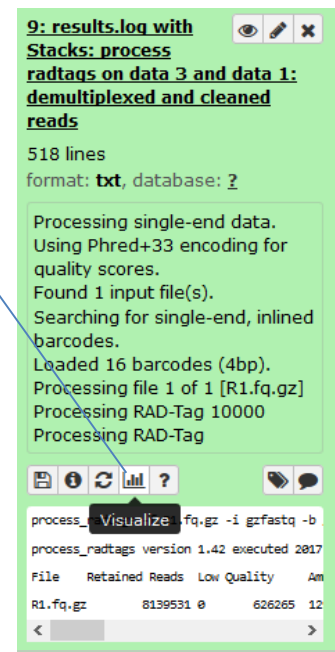
Présentation : Mode analyse



LA VISUALISATION

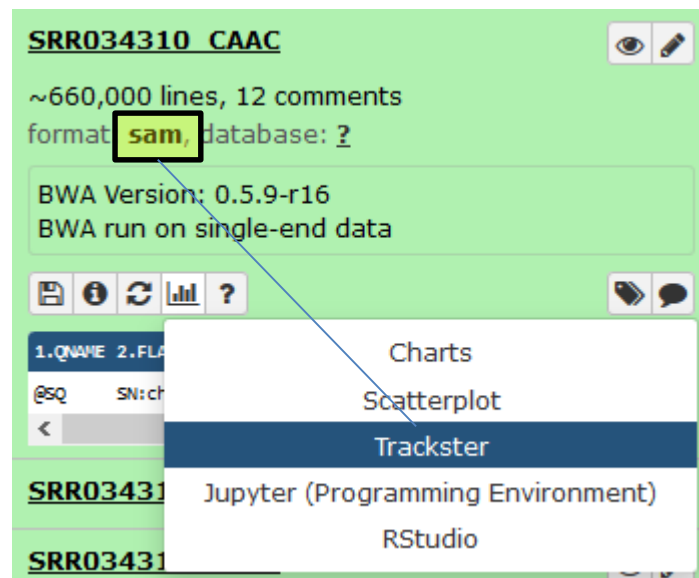
Différents outils de visualisation

- Les outils très interactifs et « lourds »
 - Il s'agit de visualisateurs associés à un format de donnée. Quand une donnée de ce format est présente dans un historique, cette visualisation est directement accessible via le bouton « **visualize** » du jeu de données
 - Ces visualisateurs sont souvent « lourds » dans le sens que ce sont des outils complets et complexes, écrits en javascript, pouvant souvent fonctionner hors Galaxy



Différents outils de visualisation

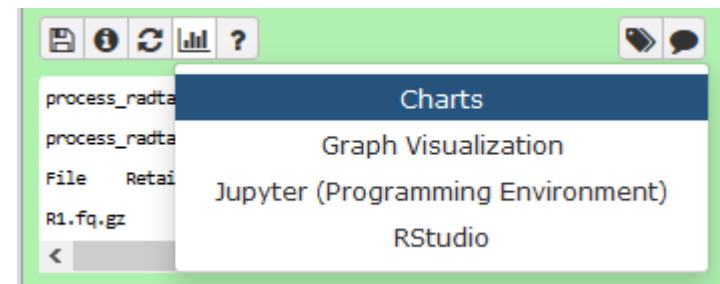
- Les outils très interactifs et « lourds »
 - Il s'agit de visualisateurs associés à un format de donnée. Quand une donnée de ce format est présente dans un historique, cette visualisation est directement accessible via le bouton « visualize » du jeu de données
 - Ces visualisateurs sont souvent « lourds » dans le sens que ce sont des outils complets et complexes, écrits en javascript, pouvant souvent fonctionner hors Galaxy



Ici, le format de fichier « sam » représente l'alignement de séquences NGS sur un génome de référence. Ce format peut être lié à un visualisateur de génome, Trackster

Différents outils de visualisation

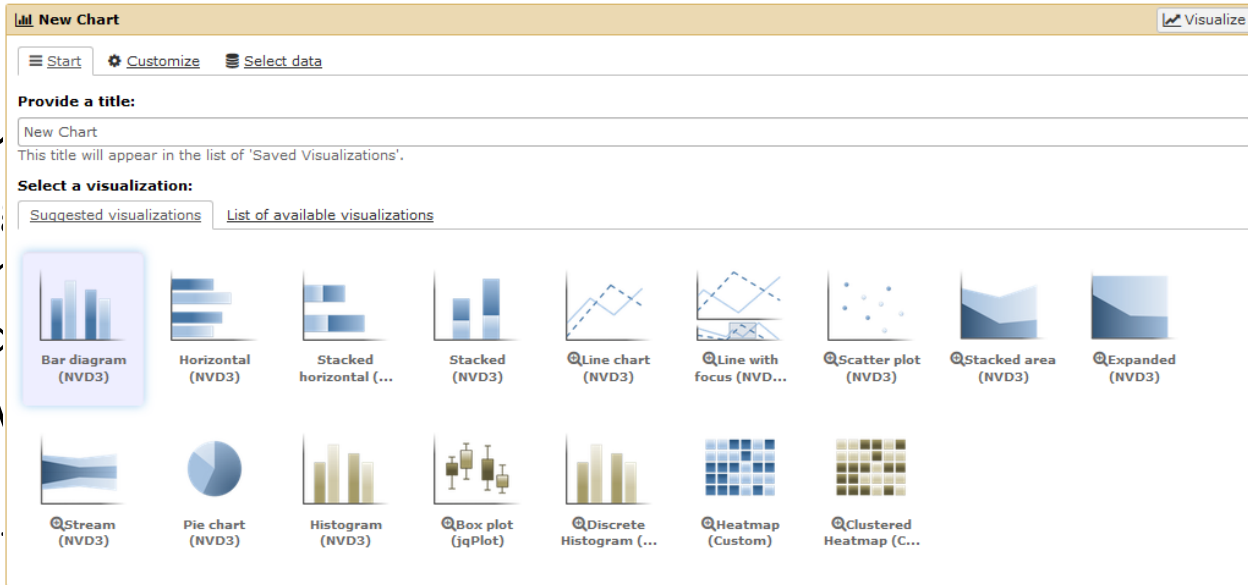
- Les outils très interactifs et « lourds »
 - Il s'agit de visualisateurs associés à un format de donnée. Quand une donnée de ce format est présente dans un historique, cette visualisation est directement accessible via le bouton « visualize » du jeu de données
 - Ces visualisateurs sont souvent « lourds » dans le sens que ce sont des outils complets et complexes, écrits en javascripts, pouvant souvent fonctionner hors Galaxy
- Les outils interactifs « légers »
 - Il s'agit de visualisateurs associés aux formats textes, notamment tabulé. Ils permettent de représenter graphiquement le contenu d'un fichier/jeu de données avec possibilité de sélectionner les colonnes à afficher et les en-têtes et autre.
 - Il s'agit des outils « Charts »



Différents outils de visualisation

- Les outils

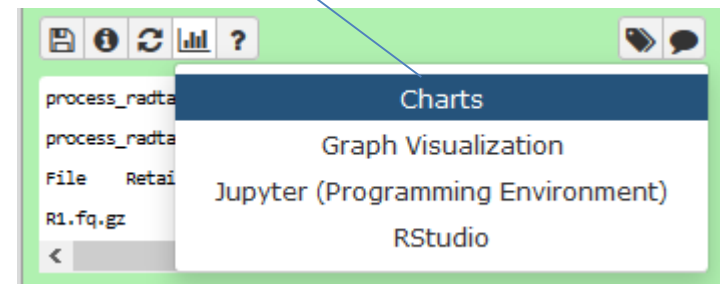
- Il s'agit de visualisateurs de données est-ce-à-dire des outils
- Ces outils sont conçus pour fonctionner avec des données



d'une
visualisation
des données
des outils

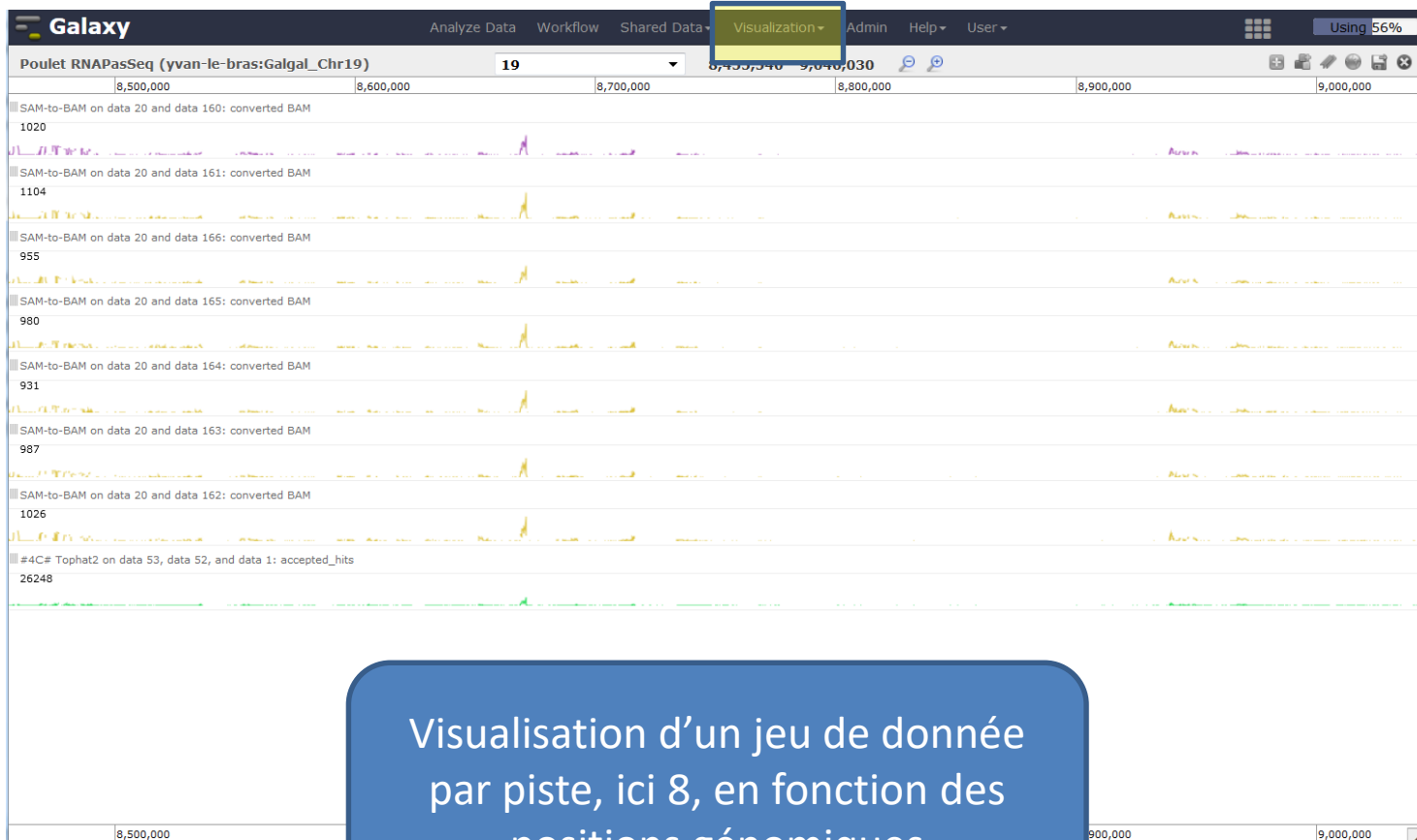
- Les outils interactifs « légers »

- Il s'agit de visualisateurs associés aux formats textes, notamment tabulé. Ils permettent de représenter graphiquement le contenu d'un fichier/jeu de données avec possibilité de sélectionner les colonnes à afficher et les en-têtes et autre.
- Il s'agit des outils « Charts »



Présentation : Mode visualisation lourd

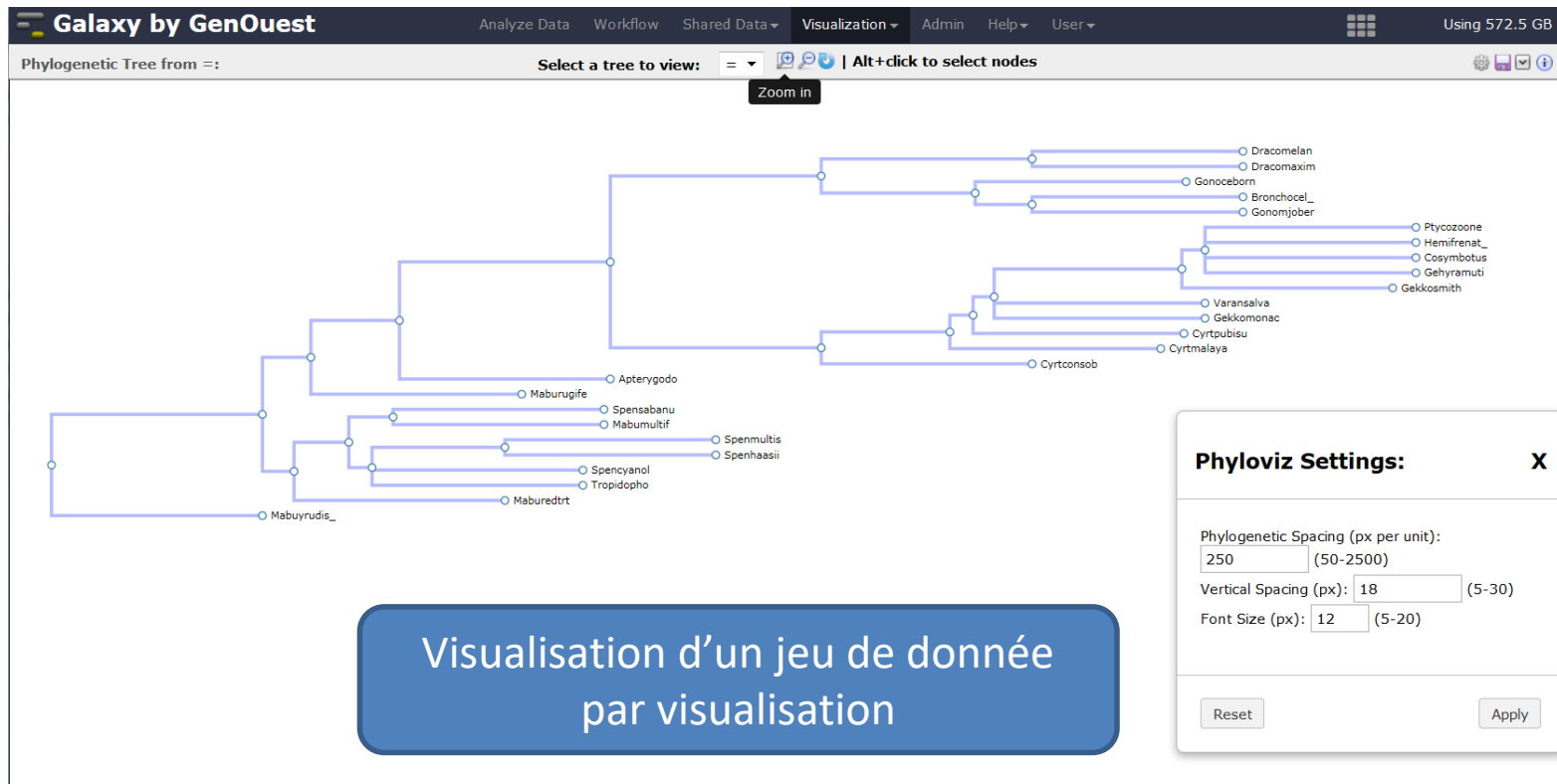
Trackster : orienté NGS et génomique



Visualisation d'un jeu de donnée
par piste, ici 8, en fonction des
positions génomiques

Présentation : Mode visualisation lourd

Phyloviz : Arbres phylogénétique



Présentation : Mode visualisation léger

Galaxy / GenOuest | Analyze Data | Workflow | Shared Data | Visualization | Admin | Help | User | Using 45%

Tools | search tools

DATA TRANSFER
[Get Data](#)
[Send Data](#)
[Archive tools](#)

SYMBIOSE
[Metagenomics benchmark](#)
[Text Manipulation](#)
[Fasta to OTUs](#)
[NGS: 454 data Manipulation](#)
[Motif](#)
[NGS: Alignment](#)
[NGS: Assembly](#)
[NGS: Genome manipulation](#)
[Systems biology](#)
[Data management](#)
[GenOuest: Mothur toolkit](#)
[KLAIST](#)

COLIBREAD
[Colib' read tool suite](#)

BIOGENOUEST COMMUNITY
[QTL Detection](#)

GENOMICS
[Alignment](#)

GENETICS
[Primer design](#)
[NGS: Contigs assembly](#)
[Population genetics](#)

PHYLOGENY
[Text Manipulation](#)
[Next Generation Quality](#)
[Phylogenetic Tree](#)
[Alignment](#)
[Blat](#)
[Phylostatistics](#)

New Chart | Draw

[Start](#) | [Configuration](#) | [Add Data](#)

Provide a chart title:
New Chart
This title will appear in the list of 'Saved Visualizations'. Charts are saved upon creation.

How many data points would you like to analyze?
[Few \(<500\)](#) | [Some \(<10k\)](#) | [Many \(>10k\)](#)

Bar diagrams

- [Regular \(NVD3\)](#)
- [Stacked \(NVD3\)](#)
- [Horizontal \(NVD3\)](#)
- [Stacked horizontal \(NVD3\)](#)

Others

- [@Line with focus \(NVD3\)](#)
- [@Line chart \(NVD3\)](#)
- [@Scatter plot \(NVD3\)](#)
- [@Heatmap \(Custom\)](#)

Area charts

- [@Regular \(NVD3\)](#)
- [@Expanded \(NVD3\)](#)
- [@Stream \(NVD3\)](#)
- [Pie chart \(NVD3\)](#)

Data processing (requires 'charts' tool from Toolshed)

- [Histogram \(NVD3\)](#)
- [@Discrete Histogram \(jqPlot\)](#)
- [@Box plot \(jqPlot\)](#)
- [@Clustered Heatmap \(Custom\)](#)

History | search datasets

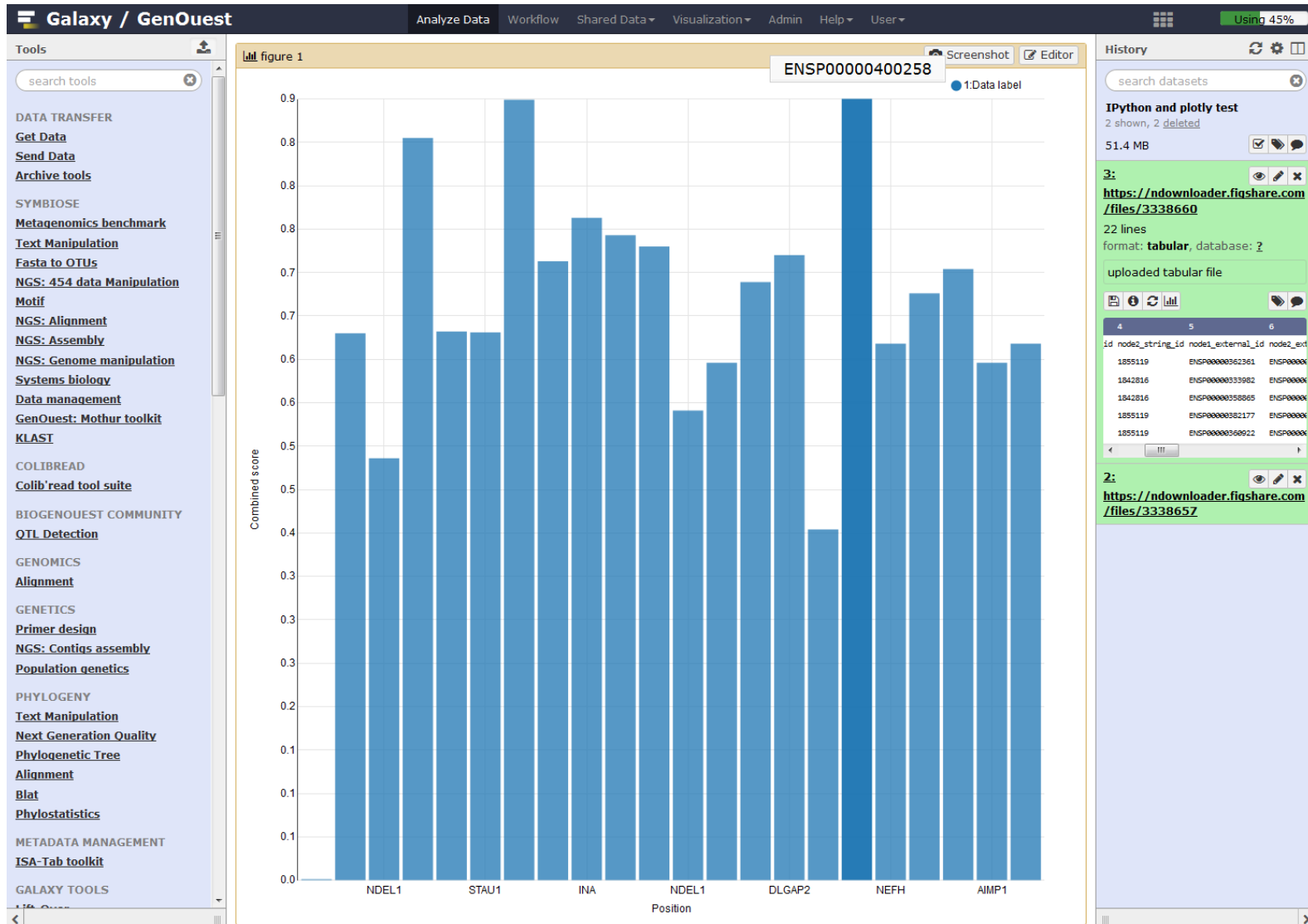
IPython and plotly test
2 shown, 2 deleted
51.4 MB
[https://ndownloader.figshare.com/files/3338660](#)
22 lines
format: **tabular**, database: ?
uploaded tabular file

Charts

1	2
node1	node2
CDK9	INA
NDEL1	NEFH
INA	NEFH
MYO5A	INA
STAU1	INA

Visualisation d'un jeu de donnée tabulé

Présentation : Mode visualisation léger



Présentation : Mode visualisation léger

Galaxy / GenQuest

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 45%

Tools

search tools

DATA TRANSFER

Get Data

Send Data

Archive tools

SYMBIOSE

Metagenomics benchmark

Text Manipulation

Fasta to OTUs

NGS: 454 data Manipulation

Motif

NGS: Alignment

NGS: Assembly

NGS: Genome manipulation

Systems biology

Data management

GenQuest: Mothur toolkit

KLAST

COLIBREAD

Colib' read tool suite

BIOGENQUEST COMMUNITY

QTL Detection

GENOMICS

Alignment

GENETICS

Primer design

NGS: Contigs assembly

Population genetics

PHYLOGENY

Text Manipulation

Next Generation Quality

Phylogenetic Tree

Alignment

Blat

Phylostatistics

Scatterplot of 'https://ndownloader.figshare.com/files/3338660'

https://ndownloader.figshare.com/files/3338660 : uploaded tabular file

Data Controls Chart Controls Chart

Save

Use the following control to change which columns are used by the chart. Click any cell from the last three rows of the table to select the column for the appropriate data. Use the 'Draw' button to render (or re-render) the chart with the current settings.

X Column: which column values will be used for the x axis of the chart.

Y Column: which column values will be used for the y axis of the chart.

ID Column: an additional column value displayed when the user hovers over a data point. It may be unique or categorical identifiers here (such as gene ids).

Columns:	1	2	3	4	5	6	7		
	node1	node2	node1_string_id	node2_string_id	node1_external_id	node2_external_id	neighborhood	fusion	cooc
	CDK9	INA	1855868	1855119	ENSP00000362361	ENSP00000358865	0.000	0.000	0.00
	NDEL1	NEFM	1851656	1842816	ENSP00000333982	ENSP00000322166	0.000	0.000	0.00
	INA	NEFM	1855119	1842816	ENSP00000358865	ENSP00000322166	0.000	0.000	0.00
	MYO5A	INA	1858783	1855119	ENSP00000382177	ENSP00000358865	0.000	0.000	0.00
	STAU1	INA	1855564	1855119	ENSP00000360922	ENSP00000358865	0.000	0.000	0.00
X Column:	X Column	X Column	X Column	X Column	X Column	X Column	X Column	X Column	X C
Y Column:	Y Column	Y Column	Y Column	Y Column	Y Column	Y Column	Y Column	Y Column	Y C
ID Column:	ID Column	ID Column	ID Column	ID Column	ID Column	ID Column	ID Column	ID Column	ID

Note: If it can be determined from the dataset's filetype that a column is not numeric, that column choice may be disabled for either the x or y axis.

Draw

Scatterplot

Visualisation d'un jeu de donnée tabulé

History

search datasets

IPython and plotly test

2 shown, 2 deleted

51.4 MB

3:

https://ndownloader.figshare.com/files/3338660

22 lines

format: tabular, database: ?

uploaded tabular file

Charts

Scatterplot

IPython

RStudio

2:

https://ndownloader.figshare.com/files/3338657

Présentation : Mode visualisation léger

Galaxy / GenOuest

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 45%

Tools

search tools

DATA TRANSFER

[Get Data](#)

[Send Data](#)

[Archive tools](#)

SYMBIOSE

[Metagenomics benchmark](#)

[Text Manipulation](#)

[Fasta to OTUs](#)

[NGS: 454 data Manipulation](#)

[Motif](#)

[NGS: Alignment](#)

[NGS: Assembly](#)

[NGS: Genome manipulation](#)

[Systems biology](#)

[Data management](#)

[GenOuest: Mothur toolkit](#)

[KLAIST](#)

COLIBREAD

[Colib' read tool suite](#)

BIOGENOUEST COMMUNITY

[QTL Detection](#)

GENOMICS

[Alignment](#)

GENETICS

[Primer design](#)

[NGS: Contigs assembly](#)

[Population genetics](#)

Scatterplot of 'https://ndownloader.figshare.com/files/3338660'

https://ndownloader.figshare.com/files/3338660 : uploaded tabular file

Data Controls

Chart Controls

Chart

Save

1 to 22 of 22

Stats

Redraw

combined score

genomic position

INA
1855119
0.763

History

search datasets

IPython and plotly test

2 shown, 2 deleted

51.4 MB

3:

<https://ndownloader.figshare.com/files/3338660>

22 lines

format: tabular, database: ?

uploaded tabular file

Charts

Scatterplot

IPython

RStudio

1 2

node1 node2

CDK9 INA

NDEL1 NEFM

INA NEFM

MMOSA INA 1858783 1855119

STAU1 INA 1855564 1855119

2:

<https://ndownloader.figshare.com/files/3338657>

Présentation : Mode visualisation léger

Galaxy / GenOuest

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 45%

Tools

search tools

DATA TRANSFER

[Get Data](#)

[Send Data](#)

[Archive tools](#)

SYMBIOSE

[Metagenomics benchmark](#)

[Text Manipulation](#)

[Fasta to OTUs](#)

[NGS: 454 data Manipulation](#)

[Motif](#)

[NGS: Alignment](#)

[NGS: Assembly](#)

[NGS: Genome manipulation](#)

[Systems biology](#)

[Data management](#)

[GenOuest: Mothur toolkit](#)

[KLAUST](#)

COLIBREAD

[Colib' read tool suite](#)

BIOGENOUEST COMMUNITY

[QTL Detection](#)

GENOMICS

[Alignment](#)

GENETICS

[Primer design](#)

[NGS: Contigs assembly](#)

[Population genetics](#)

Scatterplot of 'https://ndownloader.figshare.com/files/3338660'

https://ndownloader.figshare.com/files/3338660 : uploaded tabular file

Data Controls Chart Controls Chart

Save

1 to 22 of 22 Plot Redraw

	genomic position	combined score
min	0	0
max	1860213	0.9
sum	38956653	14.144000000000002
mean	1855078.7142857143	0.6735238095238096
median	1851656	0.541
count	21	21

History

search datasets

IPython and plotly test

2 shown, 2 deleted

51.4 MB

3:

<https://ndownloader.figshare.com/files/3338660>

22 lines

format: tabular, database: 2

uploaded tabular file

1	2	3	4
node1	node2	node1_string_id	node2_string_id
CDK9	INA	1855868	1855119
NDEL1	NEFH	1851656	1842816
INA	NEFH	1855119	1842816
MYO5A	INA	1858785	1855119
STAU1	INA	1855564	1855119

2:

<https://ndownloader.figshare.com/files/3338657>

Quelques stats

LES WORKFLOWS

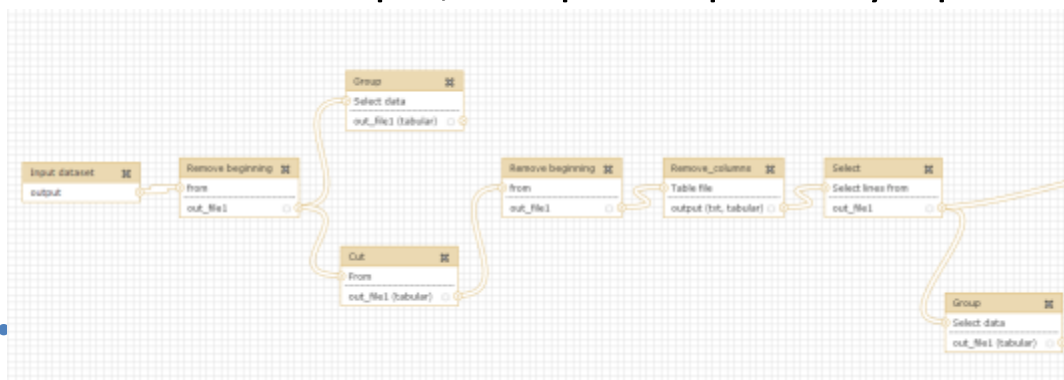
Concepts, principes, principaux outils

Principes du workflow

- Génériques
 - Automatisation des processus d'analyse (outil/composant) en les reliant dans un pipeline
 - Lancer des analyses sur des architectures matérielles complexes
 - Cluster
 - Grilles de calculs
 - Cloud
 - Formalisation du processus d'analyse
 - Enchaînement de boîtes

Le workflow sous Galaxy

- Un workflow est une extraction d'un historique analytique existant
 - L'utilisateur enchaîne ses analyses dans un historique pour aller de la donnée brute à la donnée finale analysée
 - Dans un historique, chaque étape analytique est enregistrée, avec



enchainement de boîtes analytiques avec des entrees/sorties determinees et des paramètres, donc un workflow, auquel s'ajoute les jeux de données en eux-mêmes.

Le workflow sous Galaxy

History

search datasets

Commet metagenomics datasets comparison

12 shown

6.3 MB

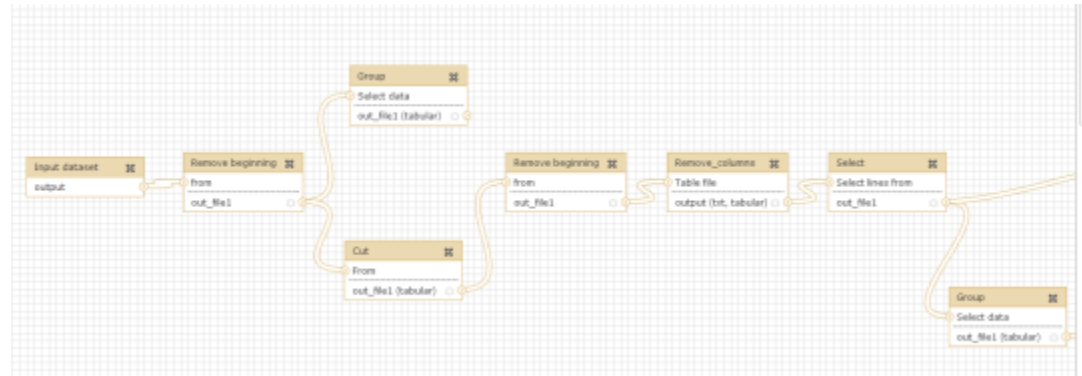
12: commet on data 3, data 2, and others: heatmap_plain.png			
11: commet on data 3, data 2, and others: heatmap_percentage.png			
10: commet on data 3, data 2, and others: heatmap_normalized.png			
9: commet on data 3, data 2, and others: matrix.zip			
8: commet on data 3, data 2, and others: dendrogram.png			
7: commet on data 3, data 2, and others: logs.zip			
6: commet on data 3, data 2, and others: vector.zip			
5: Galaxy1-[F.a].fasta			

12,000 sequences
format: **fasta**, database: ?

uploaded fasta file

```
>1
AAAAAGCATTAAGTGAACAA089A0091009TG04C00000CC0CAATCTATAATTACCA
>2
AAAAAATTTSTG0TAA9A7G049G0A0CTTTC0GCATTAACGCCATCAAACTCTGG5A
>6
AAAAACCGATTTGG0CC0GAC0CTGTCTT1G0CC0G98T1ACTCTACGCCACCAAGC98BT
>
```

4: Galaxy2-[B.f.a].fasta			
3: Galaxy3-[C.f.a].fasta			
2: Galaxy4-[D.f.a].fasta			
1: Galaxy5-[E.f.a].fasta			



De l'historique au workflow

Notion d'historique

Provenance des données

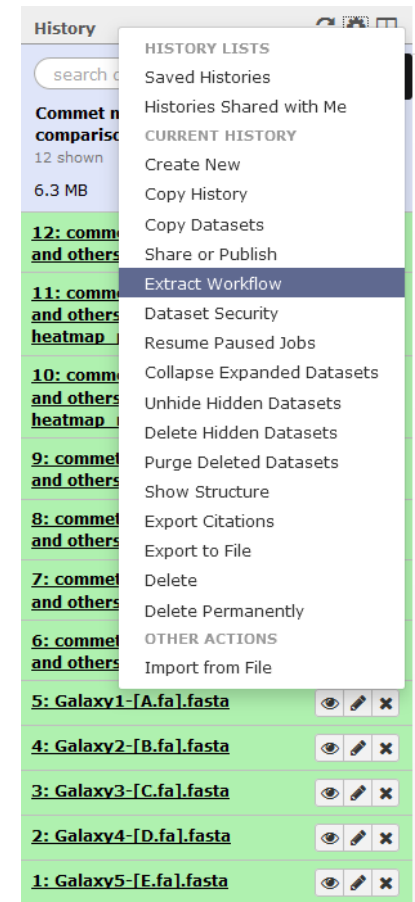


Suivi des traitements



Notion de workflow

Conversion vers un workflow



Créer un workflow

Galaxy / GenOuest

Analyze DataWorkflowShared DataVisualizationAdminHelpUser

Tools

search tools

DATA TRANSFER
[Get Data](#)
[Send Data](#)
[Archive tools](#)
SYMBIOSE
[Metagenomics benchmark](#)
[Text Manipulation](#)
[Fasta to OTUs](#)
[NGS: 454 data Manipulation](#)
[Motif](#)
[NGS: Alignment](#)
[NGS: Assembly](#)
[NGS: Genome manipulation](#)
[Systems biology](#)
[Data management](#)
[GenOuest: Mothur toolkit](#)
[KLAIST](#)
COLIBREAD
[Colib' read tool suite](#)
BIOGENOUEST COMMUNITY
[OTL Detection](#)
GENOMICS
[Alignment](#)
GENETICS
[Primer design](#)
[NGS: Contigs assembly](#)
[Population genetics](#)
PHYLOGENY
[Text Manipulation](#)
[Next Generation Quality](#)
[Phylogenetic Tree](#)
[Alignment](#)
[Blat](#)

The following list contains each tool that was run to create the datasets in your current history. Please select those that you wish to include in the workflow.
Tools which cannot be run interactively and thus cannot be incorporated into a workflow will be shown in gray.

Workflow name

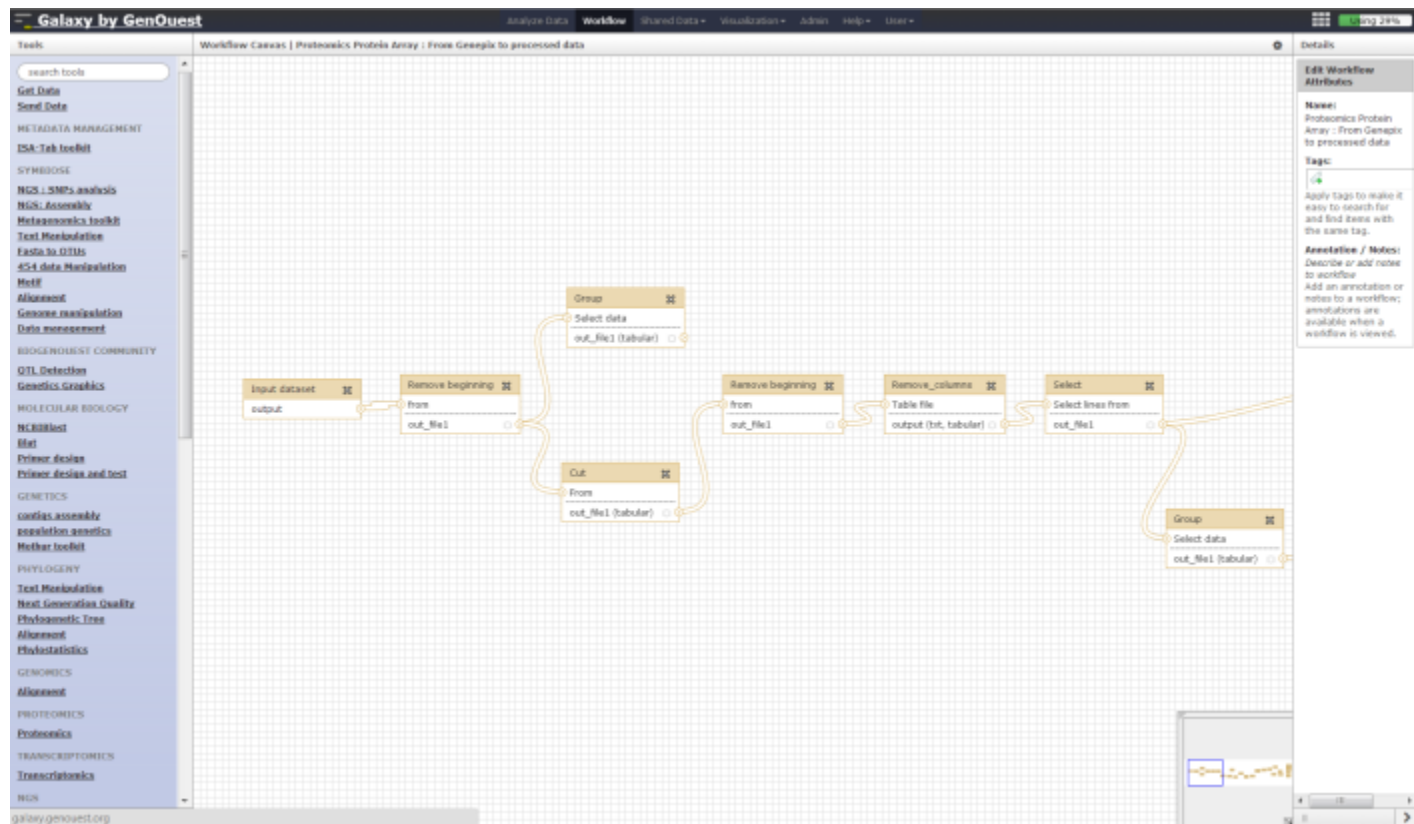
Workflow constructed from history 'Commet metagenomics c'

Create WorkflowCheck allUncheck all

Tool	History items created
Upload File This tool cannot be used in workflows	1: Galaxy5-[E.fa].fasta ☒ Treat as input dataset
Upload File This tool cannot be used in workflows	2: Galaxy4-[D.fa].fasta ☒ Treat as input dataset
Upload File This tool cannot be used in workflows	3: Galaxy3-[C.fa].fasta ☒ Treat as input dataset
Upload File This tool cannot be used in workflows	4: Galaxy2-[B.fa].fasta ☒ Treat as input dataset
Upload File This tool cannot be used in workflows	5: Galaxy1-[A.fa].fasta ☒ Treat as input dataset
commet ☒ Include "commet" in workflow	6: commet on data 3, data 2, and others: vector.zip 7: commet on data 3, data 2, and others: logs.zip 8: commet on data 3, data 2, and others: dendrogram.png 9: commet on data 3, data 2, and others: matrix.zip 10: commet on data 3, data 2, and others: heatmap_normalized.png 11: commet on data 3, data 2, and others: heatmap_percentage.png 12: commet on data 3, data 2, and others: heatmap_plain.png

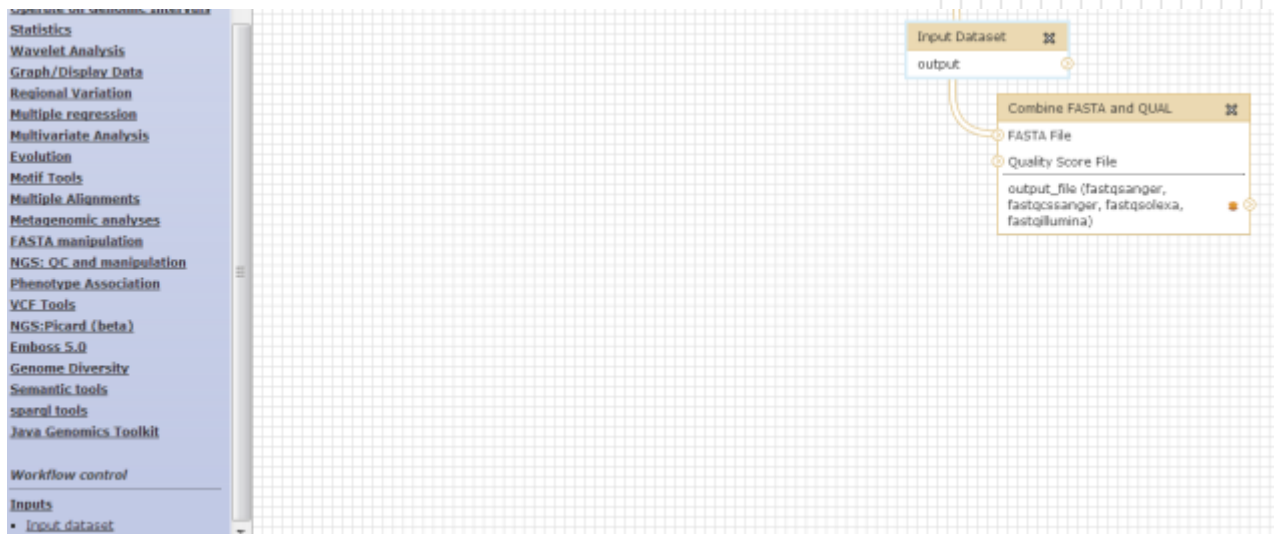
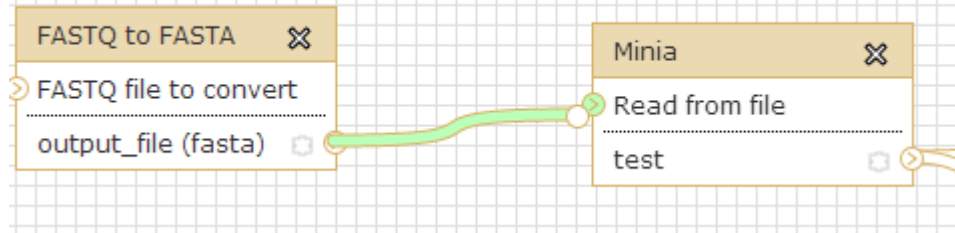
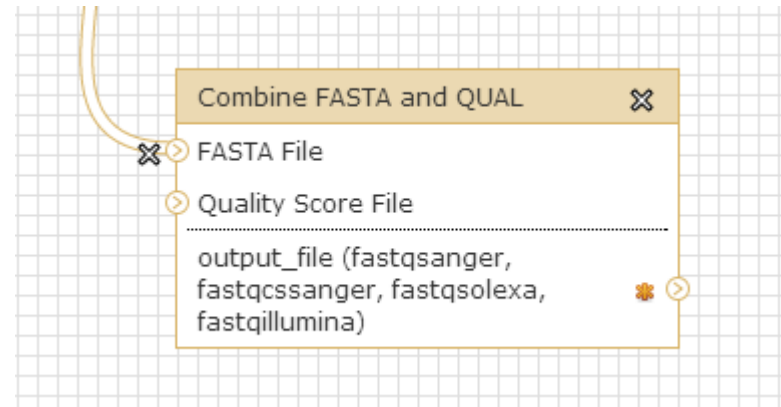
Edition d'un workflow

- Gestion de l'édition
 - Visualisation
 - Mécanisme d'intégration



Edition d'un workflow

- Gestion des composants
 - Liens entre composants
 - Ajout, modification, suppression facilitées
 - Ajout/suppression de composants



Trouver et exécuter un workflow

Galaxy / GenOuest

Analyze Data Workflow Shared Data Visualization Help User

Using 52%

Your workflows

Create new workflow Upload or import workflow

Name	# of Steps
RAD4 GTN tuto	8
Workflow constructed from history 'HUMAN exon test'	6
'Workflow constructed from history 'human exon' alternative 1'	7
Basic RNA-Seq Analysis - Differential Expression (Functional Genomics Workshop 2012) (imported from myExperiment)	4
Workflow constructed from history 'human exon'	6
Workflow constructed from history 'Tigriopus californicus transcriptome'	53
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 4: population genomics'	7
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 3: Create mini contig from PE sequences test'	12
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 1 : SNP calling.'	4
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 2: Genetic map'	102
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 5: population genomics with reference genome'	10
bwa	3
Workflow constructed from history 'Tuto Blast'	4
Workflow constructed from history 'RNAseq poulet'	8
Workflow Tuto Exon chr22 SNP	7
Workflow constructed from history 'Colib' read DiscoSNP++ : SNP Detection on raw reads'	2
Workflow constructed from history 'Tuto Galaxy exon'	7
Unnamed workflow	0

Trouver un workflow

Your workflows

Create new workflow Upload or import workflow

Name	# of Steps
RAD4 GTN tuto	8
Workflow constructed from history 'HUMAN exon test'	6
'Workflow constructed from history 'human exon' alternative 1'	7
Basic RNA-Seq Analysis - Differential Expression (Functional Genomics Workshop 2012) (imported from myExperiment)	4
Workflow constructed from history 'human exon'	6
Workflow constructed from history 'Tigriopus californicus transcriptome'	53
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 4: population genomics'	7
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 3: Create mini contig from PE sequences test'	12
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 1 : SNP calling.'	4
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 2: Genetic map'	102
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 5: population genomics with reference genome'	10
bwa	3
Workflow constructed from history 'Tuto Blast'	4
Workflow constructed from history 'RNAseq poulet'	8
Workflow Tuto Exon chr22 SNP	7
Workflow constructed from history 'Colib' read DiscoSNP++ : SNP Detection on raw reads'	2
Workflow constructed from history 'Tuto Galaxy exon'	7
Unnamed workflow	0

...

Workflows shared with you by others

Name	Owner	# of Steps
Rapsodyn - v1.23 - 10x	mathieucharles@hotmail.com	75
Rapsodyn - v1.23	mathieucharles@hotmail.com	16
'imported: Tuto SNP detection workflow pair_end' modified	alix.mas@etudiant.univ-rennes1.fr	10
SNP	fabrice.legeai@irisa.fr	5

Mes workflows

Trouver un workflow

Galaxy / GenOuest Analyze Data Workflow Shared Data Visualization Help User Using 52%

Your workflows

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

Name	# of Steps
RAD4 GTN tuto	8
Workflow constructed from history 'HUMAN exon test'	6
'Workflow constructed from history 'human exon' alternative 1'	7
Basic RNA-Seq Analysis - Differential Expression (Functional Genomics Workshop 2012) (imported from myExperiment)	4
Workflow constructed from history 'human exon'	6
Workflow constructed from history 'Tigriopus californicus transcriptome'	53
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 4: population genomics'	7
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 3: Create mini contig from PE sequences test'	12
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 1 : SNP calling.'	4
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 2: Genetic map'	102
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 5: population genomics with reference genome'	10
bwa	3
Workflow constructed from history 'Tuto Blast'	4
Workflow constructed from history 'RNAseq poulet'	8
Workflow Tuto Exon chr22 SNP	7
Workflow constructed from history 'Colib' read DiscoSNP++ : SNP Detection on raw reads'	2
Workflow constructed from history 'Tuto Galaxy exon'	7
Unnamed workflow	0

• • •

Workflows shared with you by others

Name	Owner	# of Steps
Rapsodyn - v1.23 - 10x	mathieucharles@hotmail.com	75
Rapsodyn - v1.23	mathieucharles@hotmail.com	16
'imported: Tuto SNP detection workflow pair end' modified	alix.mas@etudiant.univ-rennes1.fr	10
SNP	fabrice.legeai@irisa.fr	5

Workflows
partagés de
façon privée
avec moi

Exécuter un workflow

Galaxy / GenOuest

Analyze Data Workflow Shared Data Visualization Help User

Using 52%

Your workflows

Create new workflow Upload or import workflow

Name	# of Steps
RAD4 GTN tuto ▼	8
Workflow constructed from history 'Human exon test' ▼	6
'Workflow constructed from history 'human' ▼	7
Basic RNA-Seq Analysis - Differential Expression Workshop 2012) (imported from myExperiment) ▼	4
Workflow constructed from history 'human' ▼	6
Workflow constructed from history 'Tigriops' ▼	53
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 4: population genomics' ▼	7
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 3: Create mini contig from PE sequences test' ▼	12
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 1 : SNP calling.' ▼	4
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 2: Genetic map' ▼	102

Cliquer sur le triangle et cliquer sur « Run »