

M2 Mycologie

Outils bioinformatiques

Utilisation des serveurs Galaxy publics

Techniques bioinformatiques

Algorithmes bioinformatiques

Références bibliographiques

Utilisation des serveurs Galaxy publics

- créer un compte pour le transfert de fichiers et les notifications
- vérifier disponibilité des données partagées ("data only...")
- temps de calcul variable (queue, batch job intercurrents, événements), sensibilité aux paramètres par défaut, disponibilité des utilitaires

Authentication

Galaxy Europe

Tools
search tools

Get Data
Send Data
Collection Operations

GENERAL TEXT TOOLS
Text Manipulation
Convert Formats
Filter and Sort
Join, Subtract and Group

GENOMIC FILE MANIPULATION
Convert Formats
FASTA/FASTQ
Quality Control
SAM/BAM
BED
VCF/BCF
Nanopore

COMMON GENOMICS TOOLS
Operate on Genomic Intervals
Fetch Sequences / Alignments

GENOMICS ANALYSES
Annotation
Multiple Alignments
Assembly

Navigation: Home | Workflow | Visualize | Download outputs | Help | Authentication or Logout | Settings | Galaxy Wiki

Using RSE

Peace to Ukraine!

The list of laboratories that can host Ukrainian scientists can be found here. Galaxy Project has a number of positions at its EU and US sites. Contact us at ukraine@galaxyproject.org | Окремі надання пропозицій на роботу можна отримувати тут. Galaxy Project was supported financially by donors from various countries. Thank you to all the authors, presenters, and sponsors. Подяка всім авторам, доповідачам та спонсорам. Будь ласка, напишіть нам на ukraine@galaxyproject.org | [Окремі надання пропозицій на роботу можна отимувати тут](mailto:Окремі надання пропозицій на роботу можна отримувати тут), якщо ви хочете стати членом команди. Galaxy Project имеет открытые вакансии на своих европейских и американских сайтах. Контактуйте нас используя ukraine@galaxyproject.org.

GCC 2023 is a wrap!

Thank you to all the authors, presenters, and sponsors.

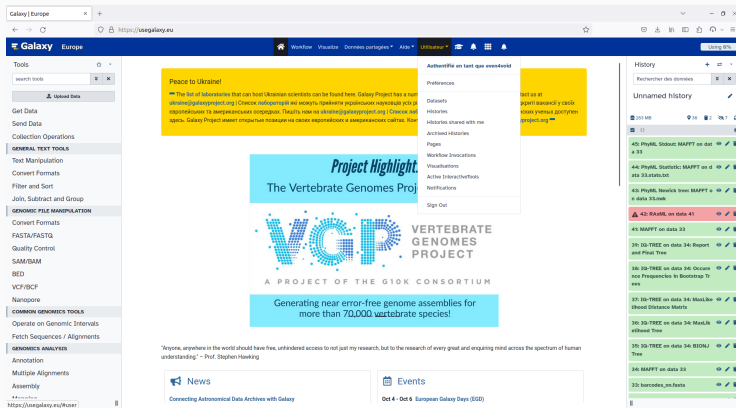
91 in-person participants	316 authors
49 virtual participants	48 talks
21 countries represented	49 posters
many koalas hugged	14 training workshops

See you next year in Brno! 🇨🇪

"Anyone, anywhere in the world should have free, unhindered access to not just my research, but to the research of every great and enquiring mind across the spectrum of human understanding." ~ Prof. Stephen Hawking






































News Events

Gestion des données partagées



The screenshot displays the Galaxy Europe web interface. The top navigation bar includes the Galaxy logo, the text "Europe", and a "Workflow" button. Below the navigation bar, the left sidebar contains a "Tools" section with a search bar and a list of tool categories: GENERAL TEXT TOOLS, GENOMIC FILE MANIPULATION, COMMON GENOMICS TOOLS, and GENOMICS ANALYSIS. The main content area features a "Project Highlight" for the Vertebrate Genomes Project (VGP), which includes a blue banner with the text "Project Highlight. The Vertebrate Genomes Project" and a logo. Below the banner, it states "Generating near error-free genome assemblies for more than 70,000 vertebrate species!". A quote from Prof. Stephen Hawking is displayed: "Anyone, anywhere in the world should have free, unbridled access to not just my research, but to the research of every great and enquiring mind across the spectrum of human understanding." - Prof. Stephen Hawking. The right sidebar shows a "History" section with a search bar and a list of unnamed history items, including "45: PhyML, Subout: MAFFT on data 33", "44: PhyML, Statistics: MAFFT on data 33, stats.txt", "43: PhyML, Newick tree: MAFFT on data 33, new", "42: RAxML, on data 41", "41: MAFFT on data 33", "39: 3D-TREE on data 34: Report and Plot Tree", "38: 3D-TREE on data 34: Occurrence Frequencies in Bootstrap Trees", "37: 10-TREE on data 34: Multi-Block Distance Matrix", "36: 3D-TREE on data 34: Multi-Block Tree", "35: 3D-TREE on data 34: BIONJ Tree", "34: MAFFT on data 33", and "33: barcodes_m.fasta".

Galaxy | Europe

Workflow | Visualize | Données partagées | Aide | **Utilisateur** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Gestion des historiques

The screenshot shows the Galaxy web interface with a 'Switch to history' dialog box open. The dialog box has a search filter and a list of history items. The first item is 'Unnamed history (Current)' with a size of 45 items and a date of 6 days ago. The second item is 'data only 2023-2024 w/p' with a size of 52 items and a date of 6 days ago. Below the list, it says '- All 2 histories loaded -'. At the bottom of the dialog, it says 'Click a history to switch to it'.

The background interface shows the Galaxy logo, a search bar, and a list of tools on the left. The main content area displays a banner for the Vertebrate Genomes Project (VGP) with the text 'Generating near error-free genome assemblies for more than 70,000 vertebrate species!'. Below the banner, there are sections for 'News' and 'Events'.

Galaxy | Europe

Tools

search tools

Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Convert Formats

Filter and Sort

Join, Subtract and Group

GENOMIC FILE MANIPULATION

Convert Formats

FASTA/FASTQ

Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

COMMON GENOMICS TOOLS

Operate on Genomic Intervals

Fetch Sequences / Alignments

GENOMICS ANALYSIS

Annotation

Multiple Alignments

Assembly

Peace to Ukraine!

The list of laboratories that are participating in the VGP project is available here. Galaxy Project website

Switch to history

Search Filter

Filter histories

Unnamed history (Current) 45 items 6 days ago

data only 2023-2024 w/p 52 items 6 days ago

- All 2 histories loaded -

Click a history to switch to it

CONTACT us at galaxy@ebi.ac.uk or galaxy@ebi.ac.uk or galaxy@ebi.ac.uk

History

Rechercher des données

Unnamed history

45: PhyML, Subout: MAFFT on data 33

44: PhyML, Statistics: MAFFT on data 33, stats.txt

43: PhyML, Newick: from MAFFT on data 33, newick

42: RAxML, on data 41

41: MAFFT on data 33

39: 10-TREE on data 34: Report and Plot Tree

38: 10-TREE on data 34: Occurrence Frequencies in Bootstrap Trees

37: 10-TREE on data 34: Multi-locus Distance Matrix

36: 10-TREE on data 34: Multi-locus Likelihood Tree

35: 10-TREE on data 34: BIONJ Tree

34: MAFFT on data 33

33: barcode_mn.fasta

News

Connecting Astronomical Data Archives with Galaxy

Events

Oct 4 - Oct 6 European Galaxy Days (EGD)

Techniques bioinformatiques

Podospira anserina

- génome 36 Mb (Fasta)
- données de séquençage 2 x 500-800 Mb (Fastq)
- 10k gènes annotés
- assemblage nouvelles souches : entre 4 et 8h (12 coeurs 3.5 GHz)
- phylogénie ITS seuls : 4h (phymI)
- phylogénie codes barres : 12 à 15h (phymI)

- serveur de calcul avec beaucoup de RAM (assemblage) et GPU (phylogénie)
- écriture de scripts shell et Python (ou R) pour les prétraitements et le développement de "workflow"
- serveur de stockage : 400 génomes ADN (+ 96 protéines) = 24 Go (en 2022)
- scripts de recherche/blast automatique (NCBI, JGI, etc.)

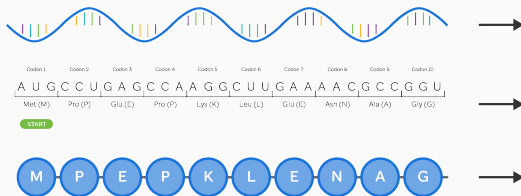
Recherche de motifs

Translation from mRNA to Protein

Mature
mRNA

Nucleotides
As Codons

Amino
Acid Sequence



© Copyright 2022 St. Jude Children's Research Hospital, a not-for-profit, section 501(c)(3)

- blast (shell ou en ligne au NCBI)
- scripts (Python, Perl, R, Bash, etc.)

Alignement de séquence

```
RLA0_METVA  --MIDAKSEHKIAPWKIEEVNALKE LLKSANVIALIDMMEVPAVLOEIRDK
RLA0_METJA  ---METKVKAHVADPKIEEVKTLKGLIKSKPVVAIVDMMDVPAPLOEIRDK
RLA0_PYRAB  -----MAHVAEWKKKKEVEELANLIKSYPVIALVDVSSMPAYPLSQMRRL
RLA0_PYRHO  -----MAHVAEWKKKKEVEELAKLIKSYPVIALVDVSSMPAYPLSQMRRL
RLA0_PYRFU  -----MAHVAEWKKKKEVEELANLIKSYPVVALVDVSSMPAYPLSQMRRL
RLA0_PYRKO  -----MAHVAEWKKKKEVEELANLIKSYPVIALVDVAGVPAYPLSKMRDK
RLA0_HALMA  MSAESERKTETIPEWKQEEVDAIVEMIESYESVGVVNIAGIPSRLODMRRD
RLA0_HALVO  MSESEVRQTEVIPQWKREEVDLVDFIESYESVGVVGAGIPSRLODSMRRE
RLA0_HALSA  MSAAEQRTTEEVPEWKQEEVDELVDLLETYDSVGVVNVGTGIPSKOLODMRRG
RLA0_THEAC  -----MKEVSQKKELVNEITRIKASRSVAIVDTAGIRTRQIQDIRGK
RLA0_THEVO  -----MRKINPKKKEIVSELAQDITKSKAVAIVDIKGVRTROMODIRAK
RLA0_PICTO  -----MTEPAQWKIDFVKNLENEINSRKVAAIVSIKGLRNNEFQKIRNS
```

- clustal
- mafft (*)
- muscle¹
- visualisateurs : jalview, seaview

¹<https://bioinformaticsreview.com/20151018/multiple-sequence-alignment/>

Mapping et assemblage (de novo)

Whole Genome Sequencing

30-60x Coverage



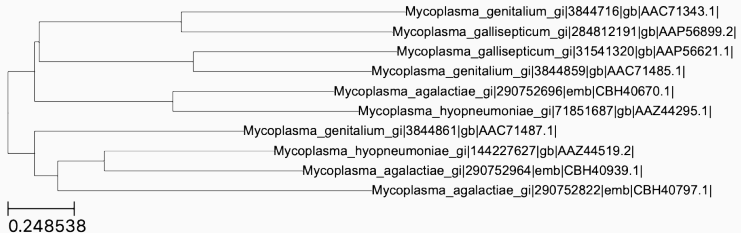
© Copyright 2022 by Jude Children's Research Hospital, a not-for-profit, section 501(c)(3)

- hisat2, tophat, bowtie2
- bwa²
- unicycler (spades) (*)
- abyss³

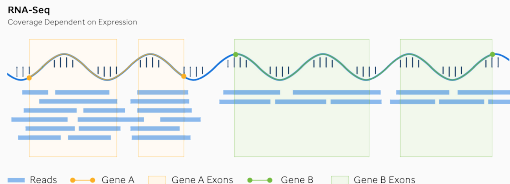
²Benchmarking short sequence mapping tools

³A biologist's guide to de novo genome assembly using next-generation sequence data

Phylogénie moléculaire

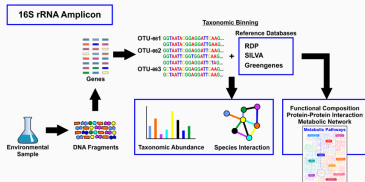


- fasttree
- IQ-TREE (*)
- RaXML
- MEGA
- NGPhylogeny
- visualisateurs : seaview (phylip), figtree, itol (payant)



© Copyright 2002-05, Jude Children's Research Hospital, a not-for-profit, section 501(c)(3)

- TopHat2 + HTSeq (ou assimilé)
- kallisto + DESeq2 (R) (*)
- Blast2Go (payant, version académique limitée)

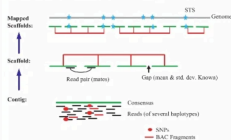


- *species*^a vs. gene-centric
- FROGS (workflow Galaxy, base de données ITS)
- Kraken (bases de données pré-existantes) (*)

^aChapter 12: Human Microbiome Analysis, PLoS Computational Biology 8(12):e1002808

Algorithmes bioinformatiques

Assemblage de génome *de novo*



Source: Venter, C. et al. 2001

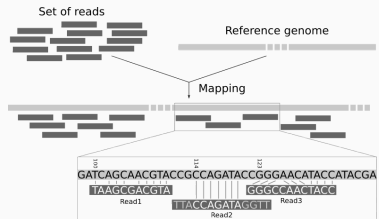
Steps for genome assembly:

- Align reads to find **overlapping regions**
- Determine a consensus sequence (or **contig**)
- **Scaffold** contigs based on read pairs and/or overlapping regions
- **Generate pseudo-molecules** based on genetic maps

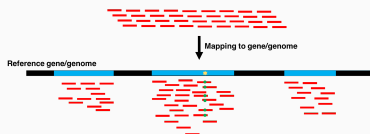
- Données : short et/ou long reads (FASTQ)
- The present and future of *de novo* whole-genome assembly

Alignement sur un génome de référence (mapping)

- Données : short reads (FASTQ), génome de référence (FASTA)
- Mapping Reads on a Genomic Sequence: An Algorithmic Overview and a Practical Comparative Analysis



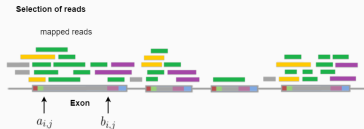
Détection de mutation (variant calling)



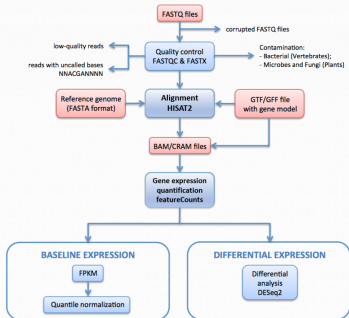
- Données : reads (FASTQ),
génom de référence (FASTA)
- Fichier VCF comprenant les
positions identifiées et les
nucléotides associés (% et
probabilité)
- Haute sensibilité aux
paramètres de filtrage (cf.
tutoriel Galaxy dans le cas
des champignons)

RNA-Seq : mapping & quantification

- Données : reads (FASTQ),
génomme de référence (FASTA)
- RPKM (reads per kilobase of
exon model per million
reads), FPKM (fragments per
kilobase of exon model per
million reads mapped) : prise
en compte de la longueur des
gènes et de la taille de la
bibliothèque
- Systematic comparison and
assessment of RNA-seq
procedures for gene
expression quantitative
analysis

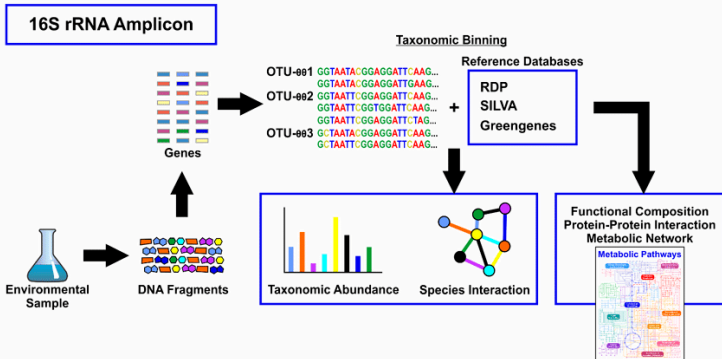


RNA-Seq : analyse différentielle

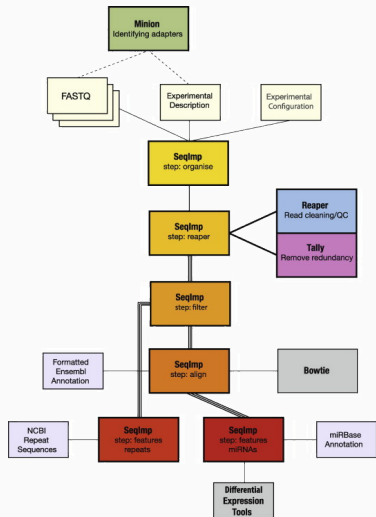


- Données : RPKM ou FPKM
- Approche fréquentiste ou bayésienne pour décider si les données de comptage moyennées sur les réplicats techniques et normalisées pour chaque réplicat biologique sont dûes au hasard ou non (gène sur- ou sous-exprimé par analyse de contraste sur condition de référence).

Métagénomique : Principe général



- Utilisation de base de données pré-définies, que l'on peut augmenter avec des souches de référence
- Mise en oeuvre rapide et rapport importable dans les suites d'analyses statistiques



Références bibliographiques

1. Ernesto Picardi (2021) *RNA Bioinformatics*, Springer
2. Ziheng Yang & Bruce Rannala (2012) 'Molecular phylogenetics: principles and practice', *Nature Reviews Genetics* 13, 303–314
3. Scot A. Kelchner & Michael A. Thomas (2006) 'Model Use in Phylogenetics: Nine Key Questions', *TRENDS in Ecology and Evolution* 22, 87–94
4. Mostafa M. Abbas, Qutaibah M. Malluhi & Ponnuraman Balkrishnan (2014) 'Assessment of de novo assemblers for draft genomes: a case study with fungal genomes', *BMC Genomics* 15, 1–12
5. Bo Li, Victor Ruotti, Ron M. Stewart, James A. Thomson & Colin N. Dewey (2010) 'RNA-Seq gene expression estimation with read mapping uncertainty', *Bioinformatics* 26, 493–500

Source principale des illustrations : <https://learngenomics.dev/>,
<https://is.gd/xRcxSR>