

# **M2 Mycologie**

Outils bioinformatiques

- préférer le serveur Europe : <https://usegalaxy.eu>
- 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

### *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.)

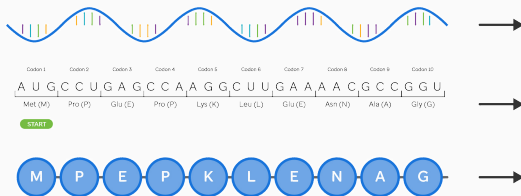
# Recherche de motifs

## 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 --MIDAKSEHKTIAPWKIEEVNALKELLKSANVIALIDMMEVPAVLOEIRDK
RLA0_METJA ---METKVKAHVADPKIEEVKTLKGLIKSKPVVAIVDMMDVPAPLOEIRDK
RLA0_PYRAB -----MAHVAEWKKKKEVEELANLIKSYPVIALVDVSSMPAYPLSQMRRL
RLA0_PYRHO -----MAHVAEWKKKKEVEELAKLIKSYPVIALVDVSSMPAYPLSQMRRL
RLA0_PYRFU -----MAHVAEWKKKKEVEELANLIKSYPVVALVDVSSMPAYPLSQMRRL
RLA0_PYRKO -----MAHVAEWKKKKEVEELANLIKSYPVIALVDVAGVPAYPLSKMRDK
RLA0_HALMA MSAESERKTETIPEWKQEEVDAIVEMIESYESVGVVNIAGIPSRLODMRRD
RLA0_HALVO MSESEVRQTEVIPQWKREEVDLVDFIESYESVGVVGAGIPSRLOSMRRE
RLA0_HALSA MSAEEQRTTEEVPEWKQREVAELVDLLETYDSVGVVNVGTGIPSKOLODMRRG
RLA0_THEAC -----MKEVSQKKELVNEITRIKASRSVAIVDTAGIRTRQIQDIRGK
RLA0_THEVO -----MRKINPKKKEIVSELAQDITKSKAVAIVDIKGVRTROMODIRAK
RLA0_PICTO -----MTEPAQWKIDFVKNLENEINSRKVAAIVSIKGLRNNEFQKIRNS
```

- clustal
- mafft (\*)
- muscle<sup>1</sup>
- visualisateurs : jalview, seaview

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

# Mapping et assemblage (de novo)

## Whole Genome Sequencing

30-60x Coverage



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

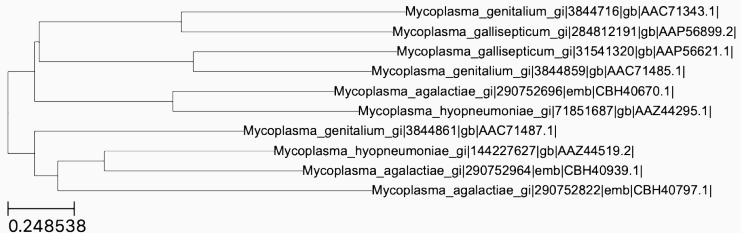
- bowtie2
- bwa<sup>2</sup>
- unicycler (spades) (\*)
- abyss<sup>3</sup>

---

<sup>2</sup>Benchmarking short sequence mapping tools

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

# Phylogénie moléculaire

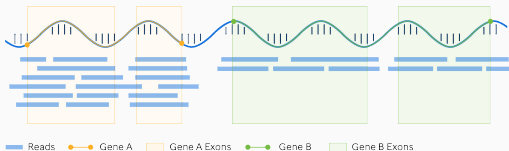


- fasttree
- IQ-TREE (\*)
- RaXML
- MEGA
- NGPhylogeny
- visualisateurs : seaview (phylip), figtree, itol (payant)



## RNA-Seq

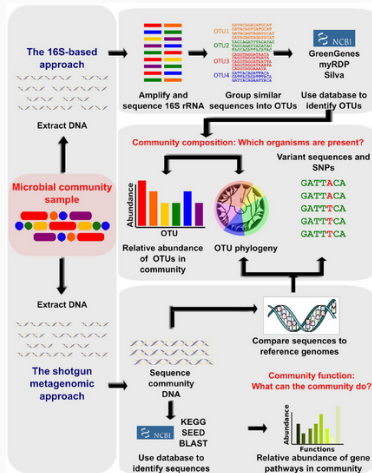
Coverage Dependent on Expression



© Copyright 2002-05, Jude Children's Research Hospital, a not-for-profit, we have 5202 (g/f)

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

# Métagénomique



- *species*<sup>a</sup> vs. gene-centric
- FROGS (workflow Galaxy, base de données ITS)
- Kraken (bases de données pré-existantes) (\*)

<sup>a</sup>Chapter 12: Human Microbiome Analysis, PLoS Computational Biology 8(12):e1002808

## 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/>