



UNIVERSITE DE JIJE



Faculté des sciences de la nature et de la vie

Master I Biochimie  
Groupe :02

### **Rapport TP :**

### **Logiciels Libres et Open Source**

- **Préparé par:**

-KAHLESENANE Roumaissa

-GUEHAM Manel

- **Année universitaire: 2025-2026**

## Introduction

**Les logiciels libres et open source (FOSS) sont des logiciels disponibles sous une licence qui donne aux utilisateurs le droit d'utiliser, de partager, de modifier et de distribuer le logiciel -modifié ou non à tous et fournit les moyens d'exercer ces droits en utilisant le code source du logiciel.(1)**

## PARTIE 1 : Étude théorique d'un outil EMBOSS

### 1-Présentation générale de l'outil EMBOSS :

**EMBOSS est un logiciel gratuit en open source de l'Institut européen de bio-informatique (EBI) dont l'acronyme signifie European Molecular Biology Open Software Suite. Il a été développé à l'intention des utilisateurs dans les domaines de la biologie moléculaire et de la bio-informatique.(2)**



### 2-Fonctionnalités principales :

- Conversion et manipulation de séquences
- Alignement multiple
- Analyse de motifs génétiques
- Visualisation simples
- Format compatible avec plusieurs bases de données(3)

### 3-Aspects techniques :

- Langage de programmation : C, pour un traitement rapide et une bonne gestion de la mémoire.
- Architecture modulaire: chaque outil est autonome et peut être combiné dans des workflows.
- Gestion des données: compatible avec de nombreux formats (FASTA, GenBank, EMBL, GFF...)
- Bibliothèques internes: gèrent les entrées/sorties, les manipulations de séquences et les algorithmes bioinformatiques.

**Algorithmes principaux :**

Smith-Waterman pour alignements locaux (water)

Needleman-Wunsch pour alignements globaux (needle)

Méthodes semi-empiriques pour prédiction d'antigènes (antigenic). (4)

#### **4-Points forts :**

**EMBOSS est hautement portable, gratuit et bien documenté, idéal pour les environnements Unix sans interface graphique. Sa richesse en applications (plus de 200) et son intégration avec des outils tiers en font un standard académique fiable pour l'analyse de séquences. Il offre une flexibilité pour les scripts personnalisés et une communauté active.(5)**

#### **5-Limites et points faibles :**

**L'interface en ligne de commande peut rebuter les débutants sans expérience Unix, manquant d'interface graphique moderne. Certains outils sont spécialisés mais moins performants que des alternatives récentes comme Biopython pour des tâches complexes. La maintenance semble ralentie, avec des versions datant (ex.: 6.6.0), limitant les mises à jour pour les nouvelles technologies.(6)**

#### **6- Conclusion :**

**EMBOSS est une suite incontournable pour l'analyse bioinformatique classique des séquences biologiques.**

---

## **PARTIE 2 : Étude pratique ; exploration de Zenodo**

### **1. Présentation de Zenodo :**

**Zenodo est un entrepôt pluridisciplinaire lancé en 2013 pour soutenir la politique d'Open Access et d'Open Data de la Commission européenne.** Développé et hébergé par le CERN, Zenodo est un logiciel open source (code open source, basé sur le logiciel Invenio, lui-même open source). Zenodo est utilisable par toute la communauté scientifique pour la préservation et le partage des résultats de la recherche pour lesquels il n'existe pas de dépôt institutionnel, disciplinaire ou thématique.

#### **➤ Objectifs de la plateforme :**

- Sécurisé: stockez vos données de recherche en toute sécurité pour l'avenir dans le centre de données du CERN, aussi longtemps que le CERN existera.
- Cité: attribuez un identifiant d'objet numérique (DOI) à chaque jeu de données de recherche, afin de le rendre citable et traçable.
- Conditions d'accès et de réutilisation: indiquez clairement le niveau d'accès (ouvert, restreint, sous embargo) et la licence.
- Gestion des versions: mettez facilement à jour votre jeu de données grâce à la fonction de gestion des versions.
- Intégration GitHub: préservez facilement votre dépôt GitHub dans Zenodo et rendez-le citable. (7)

#### **➤ Types de contenus hébergés :**

**Zenodo, plateforme généraliste hébergée par le CERN, accepte tout type de contenu numérique lié à la recherche, toutes disciplines confondues, sans restriction de format.**

**L'entrepôt est ouvert à une large diversité de produits de recherche:**

- Publications scientifiques:** Zenodo peut être utilisé comme une archive ouverte pour la conservation et la diffusion des publications (c'est peu le cas en France car l'archive nationale HAL est privilégiée et soutenue par le ministère de la recherche).
- Jeux de données:** De par sa facilité d'utilisation, Zenodo est souvent utilisé pour partager des données associées à une publication (on vous explique comme faire ici).
- Logiciels:** Zenodo peut notamment être utilisé pour rendre son code facilement citable (on vous explique comment faire ici).
- Autres contenus scientifiques:** présentations, posters, images, vidéos, cours...

➤ **Intérêt de Zenodo pour la science ouverte et la recherche en science de la nature et de la vie :**

**Zenodo est un pilier de la science ouverte, augmentant l'efficacité et l'impact de la recherche en sciences de la nature et de la vie.**

✓ **Intérêts de Zenodo pour la Science Ouverte:**

- Accessibilité et visibilité : Il permet le dépôt de tous les produits de la recherche (articles, données, logiciels, images) dans tous les domaines.
- Identifiants pérennes (DOI): Chaque dépôt reçoit un DOI, facilitant la citation directe et l'attribution de crédit aux auteurs.

**-Respect des principes FAIR : Les données sont facilement trouvables et réutilisables, ce qui limite la duplication des recherches.**

**-Gestion des accès: Il offre une flexibilité totale: accès ouvert (Open Access), accès restreint, accès sur demande ou sous embargo.**

**-Environnement non commercial: Développé au CERN avec le soutien de l'Union Européenne, il garantit la pérennité des données.**

✓ **Intérêts pour les Sciences de la Nature et de la Vie :**

**-Partage de jeux de données massifs: Essentiel pour le stockage de données expérimentales, génomiques, ou d'imagerie qui nécessitent une visibilité accrue.**

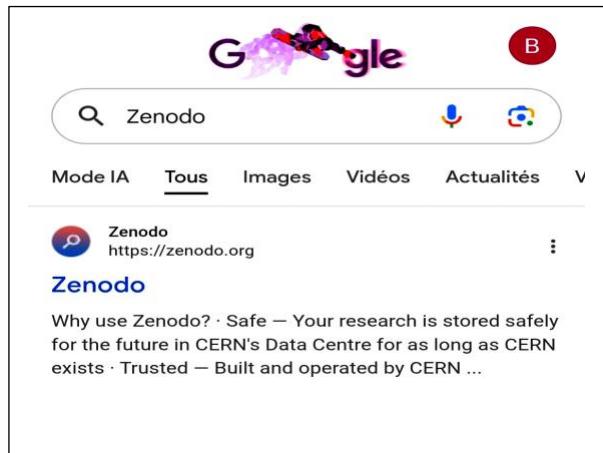
**-Reproductibilité: Il permet de lier les publications aux codes et données brutes, répondant ainsi à la crise de la reproductibilité en science.**

**-Collaboration interdisciplinaire : Facilite le partage entre équipes internationales, notamment pour les études sur l'environnement et les données in situ.**

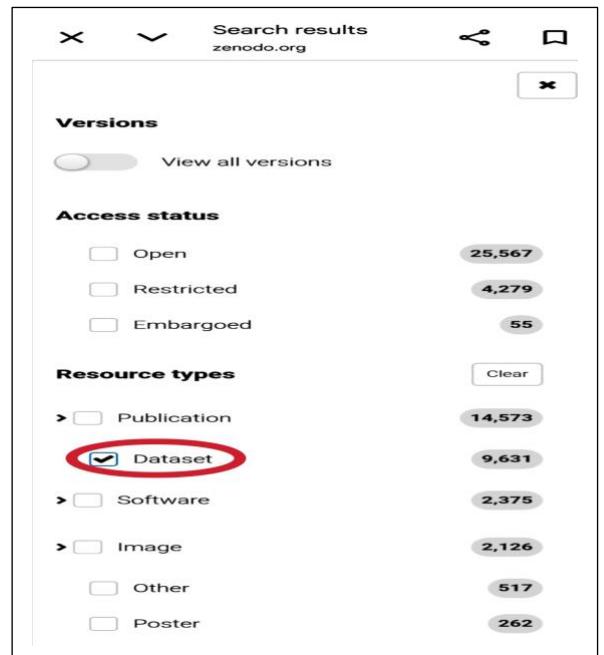
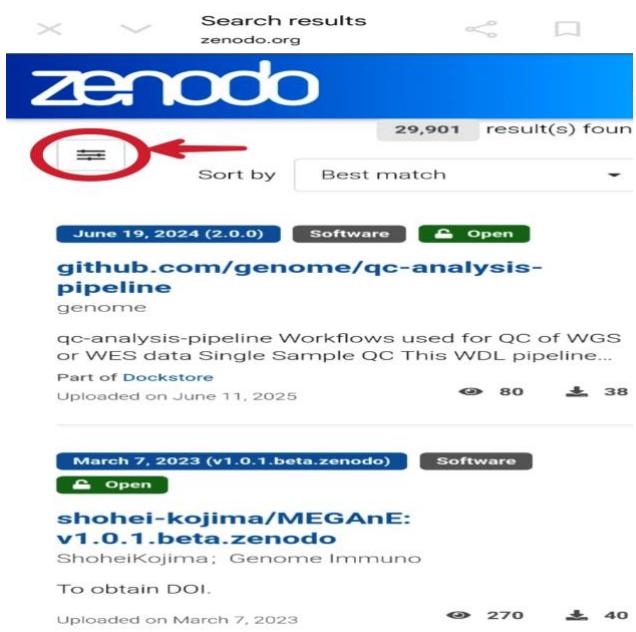
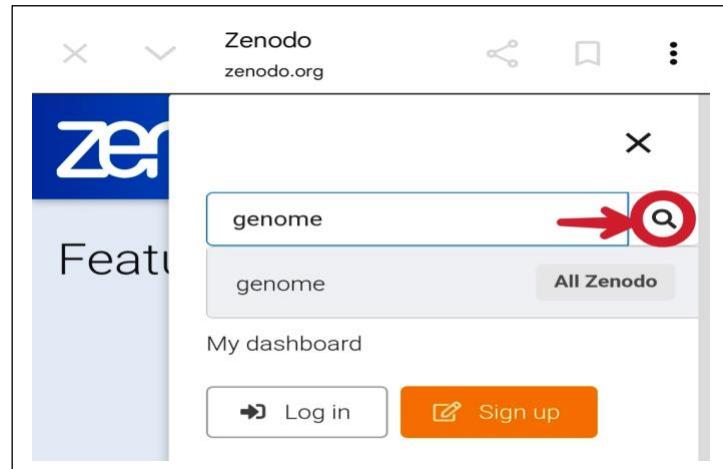
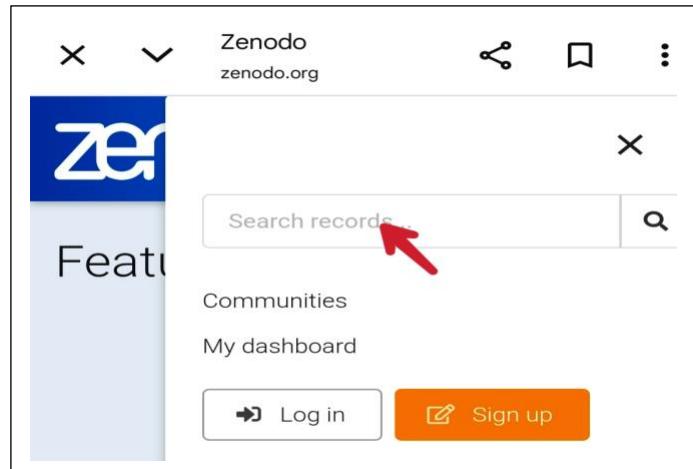
**-Compatibilité avec les financements: Conforme aux exigences des bailleurs de fonds (comme Horizon Europe) qui imposent l'ouverture des données.**

## 2. Description des étapes réalisées

### 1-Visiter la plateforme Zenodo



### 2-Effectuer une recherche de dataset



### 3-Sélectionner un dataset pertinent et Télécharger le dataset choisi.

Search results zenodo.org

May 18, 2012 (v1) Dataset Open

Data from: **Butterfly genome reveals promiscuous exchange of mimicry adaptations among species**  
Consortium, The Heliconius Genome  
The evolutionary importance of hybridization and introgression has long been debated. Hybrids are...  
Part of Dryad  
Uploaded on June 16, 2021

June 26, 2024 (v1) Dataset Open

Source data for publication "A multimodal atlas of hepatocellular carcinoma reveals convergent evolutionary paths and 'bad apple' effect on clinical trajectory" in Journal of Hepatology.  
Genome Institute of Singapore ROR  
Processed genomic and transcriptomic data for the publication...  
Uploaded on June 30, 2024

Files (73.8 MB)

source\_data.zip  
md5:1126e860774f7d08b113eef03d31cf35 ?

73.8 MB

Preview Download

### 4-Récupérer les métadonnées associées au dataset

Export

Dublin Core XML Export

DataCite JSON

DataCite XML

Dublin Core XML

Modified June 11, 2024

Export

Dublin Core XML

Export

### 6-Extraire et présenter le maximum d'informations disponibles

12548543.xml

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<oai_dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
<dc:creator>Genome Institute of Singapore</dc:creator>
<dc:date>2024-06-26</dc:date>
<dc:description>&lt;p&gt;Processed genomic and transcriptomic data for the publication &lt;a href="https://doi.org/10.1016/j.jhep.2024.05.017"&gt;https://doi.org/10.1016/j.jhep.2024.05.017&lt;/a&gt;.&lt;/p&gt;
&lt;p&gt;cnv_segmentation.tsv: CNV segmentation file from Sequenza.&lt;/p&gt; &lt;p&gt;cnv_arm.tsv: Significant arm level CNV events called by GISTIC, from broad_values_by_arm.txt file.&lt;/p&gt; &lt;p&gt;cnv_gene.tsv: Gene level CNV events called by GISTIC, from all_threshold_by_genes.txt file.&lt;/p&gt; &lt;p&gt;RNA_raw_counts.tsv: Raw RNA-seq read counts from featureCounts.&lt;/p&gt; &lt;p&gt;snp_indel.tsv: All SNV and Indel called with annotation from Functator.&lt;/p&gt; &lt;p&gt;snv_indel.tsv: All SNV and Indel called with annotation from Functator.&lt;/p&gt; &lt;p&gt;snv_indel.tsv: All SNV and Indel called with annotation from Functator.&lt;/p&gt;</dc:description>
<dc:identifier>https://doi.org/10.5281/zenodo.12548543</dc:identifier>
<dc:identifier>oai:zenodo.org:12548543</dc:identifier>
<dc:publisher>Zenodo</dc:publisher>
<dc:relation>https://doi.org/10.1016/j.jhep.2024.05.017</dc:relation>
<dc:relation>https://doi.org/10.5281/zenodo.12548542</dc:relation>
<dc:rights>info:eu-repo/semantics/openAccess</dc:rights>
<dc:rights>Creative Commons Attribution 4.0 International</dc:rights>
<dc:rights>https://creativecommons.org/licenses/by/4.0/legalcode</dc:rights>
<dc:title>Source data for publication "A multimodal atlas of hepatocellular carcinoma reveals convergent evolutionary paths and 'bad apple' effect on clinical trajectory" in Journal of Hepatology</dc:title>
<dc:type>info:eu-repo/semantics/other</dc:type>
</oai_dc:dc>
```

### Shéma représentative structurée des métadonnées récupérées

### **3-Métadonnées du dataset**

Élément Dublin Core	Valeur extraite
Title	Source data for publication “A multimodal atlas of hepatocellular carcinoma reveals convergent evolutionary paths and ‘bad apple’ effect on clinical trajectory” in <i>Journal of Hepatology</i>
Creator	Genome Institute of Singapore (Hosting institution)
Description	Processed genomic and transcriptomic data supporting the associated <i>Journal of Hepatology</i> publication (incluant CNV segmentation, CNV arm-level calls, gene-level CNV calls, RNA-seq raw counts, SNV/Indel calls).
Publisher	Zenodo
Date Published	26 June 2024 ; Created : 30 June 2024 ; Modified : 30 June 2024
Type	Dataset
Subject	Hepatocellular carcinoma ; intra-tumoral heterogeneity ; multi-region tumor sampling ; tumor evolution ; RNA-seq ; CNV ; SNV/Indel ; genomic & transcriptomic profiles
Format	ZIP archive (source_data.zip – 73.8 MB ; total data volume ~3.0 GB) ; file types include TSVs for genomic & transcriptomic data
Identifier	DOI : 10.5281/zenodo.12548543
Relation	Supplement to the associated <i>Journal of Hepatology</i> article
Coverage	Multi-omics molecular data from hepatocellular carcinoma samples
Rights	Creative Commons Attribution 4.0 International (CC-BY-4.0)

## **PARTIE 3 : BONUS**

<https://github.com/Roumaissa4180>

### **Références :**

- 1-Sources describing free and open-source software as an umbrella term encompassing both free software and open source software.
- 2- Peter Rice, Ian Longden et Alan Bleasby, <<EMBOSS: The European Molecular Biology Open Software Suite >>, Trends in Genetics, vol. 16, n° 6, juin 2000, p. 276-277
- 3- Site Scribd
- 4-Smith, T.F., & Waterman, M.S. (1981). Identification of common molecular subsequences. J Mol Biol, 147(1), 195-197.
- 5-wikipedia
- 6-EMBOSS: The European Molecular Biology Open Software Suite (2000) Rice, P. Longden, I. and Bleasby A.Trends in Genetics 16, (6) pp276-277.
- 7-2026 Universiteit Gent