



الجمهورية الجزائرية الديمقراطية الشعبية  
وزارة التعليم العالي والبحث العلمي  
République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



جامعة محمد الصديق بن يحيى جيجل

Faculté des Sciences de la Nature et de la Vie

كلية علوم الطبيعة والحياة

Département : Biologie Moléculaire et Cellulaire

قسم : البيولوجيا الجزيئية والخلوية

Filière : Sciences biologiques

المجال : العلوم البيولوجية

Spécialité : Biochimie appliquée

التخصص : الكيمياء الحيوية التطبيقية

## TP : Logiciels Libres et Open Source

Réalisé et mis en page par :

- Fennour Imane
- Zenanra Amani
- Boukhenoufa Hadil

Enseignante : Bensalem Amel

Groupe : 02

Module : Logiciels Libres et Open Source



Année Universitaire 2025 - 2026

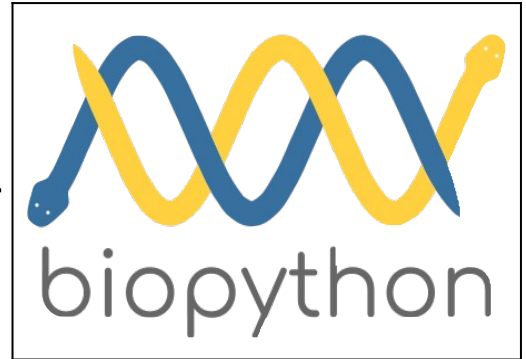
## PARTIE 1 – Étude théorique d'un outil

### 1. Présentation générale de l'outil

Biopython est une bibliothèque **open source** développée en langage **Python**, dédiée à l'analyse et au traitement des données biologiques. Elle est principalement utilisée en **bio-informatique** pour manipuler des séquences biologiques (ADN, ARN, protéines), accéder aux bases de données biologiques et effectuer des analyses computationnelles.

Biopython est développée par une **communauté internationale** de chercheurs et d'enseignants et est largement utilisée dans la recherche et l'enseignement en sciences de la vie.

L'outil a été développé par une **communauté active**, reconnu pour sa **simplicité**, sa **lisibilité** et sa **richesse** en bibliothèques scientifiques. Biopython est maintenu dans un **dépôt public**, et sa documentation est **accessible en ligne**, facilitant ainsi son adoption et son appropriation par les chercheurs et étudiants

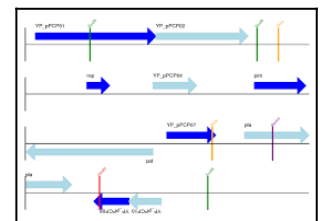
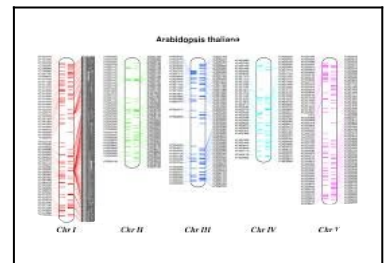
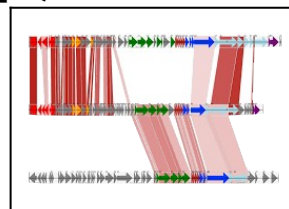


*Figure 1 : Logo officiel du projet Biopython.*

### 2. Fonctionnalités principales

Biopython offre un large ensemble de fonctionnalités, parmi lesquelles :

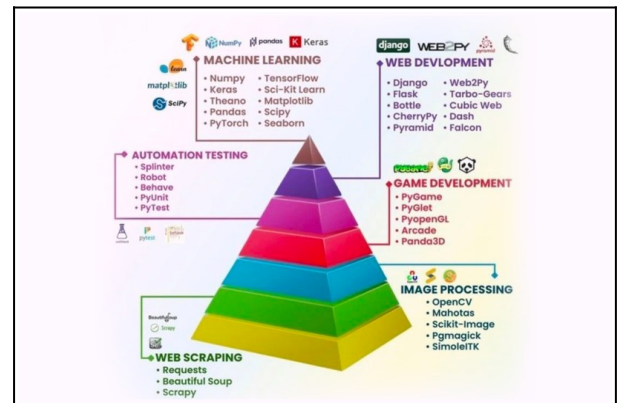
- Lecture, écriture et manipulation de séquences biologiques (FASTA, GenBank, EMBL)
- Analyse de séquences d'ADN, d'ARN et de protéines
- Alignement de séquences (pairwise et multiple)
- Accès aux bases de données biologiques en ligne (NCBI, GenBank)
- Analyse phylogénétique
- Manipulation de structures tridimensionnelles de protéines (PDB)
- Traitement des données de séquençage haut débit



*Figure 2 : Exemples de manipulation de séquences biologiques à l'aide de Biopython.*

### 3. Aspects techniques

- Langage de programmation : **Python**
- Licence : **Open source (Biopython License)**
- Systèmes compatibles : Windows, Linux, macOS
- Installation via gestionnaire de paquets :  
pip install biopython
- Compatible avec d'autres bibliothèques scientifiques (NumPy, Pandas, SciPy)



*Figure 3 : Architecture générale et intégration de Biopython avec l'écosystème Python.*

### 4. Points forts

Biopython présente plusieurs avantages importants :

#### Facilité d'utilisation

- Grâce à Python, la syntaxe de Biopython est claire et intuitive, facilitant l'apprentissage même pour des utilisateurs débutants en programmation.

#### Richesse des fonctionnalités

- La bibliothèque couvre de nombreux aspects de la bioinformatique, ce qui la rend utilisable pour des applications variées allant du traitement de séquences à l'analyse structurale.

#### Communauté active

- Biopython est soutenu par une communauté d'utilisateurs et de contributeurs qui enrichit régulièrement son code, corrige des bugs et documente de nouveaux modules

#### Intégration

- L'intégration fluide avec d'autres bibliothèques Python permet de construire des workflows puissants et personnalisés, adaptés à des projets de recherche complexes.

## 5. Limites et points faibles

- Nécessite des connaissances de base en programmation Python
- Moins adapté aux utilisateurs préférant une interface graphique
- Performances parfois limitées pour l'analyse de très grands volumes de données
- Certaines analyses complexes nécessitent des outils externes complémentaires

## 6. Conclusion

**Biopython** est une bibliothèque **puissante, flexible** et largement utilisée dans le domaine de la **bioinformatique**. Son intégration avec Python en fait un outil accessible autant aux **débutants** qu'aux **chercheurs expérimentés**. Elle permet d'automatiser et d'accélérer l'analyse de données biologiques à travers une grande variété de fonctionnalités, tout en étant soutenue par une communauté active. Malgré quelques limitations en matière de **performance** ou de **fonctionnalités** très spécifiques, Biopython demeure un outil de référence dans les **workflows** **bioinformatiques modernes**.

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

### 1. Présentation de Zenodo

Zenodo est une plateforme en ligne de dépôt et de partage de productions scientifiques, créée en 2013 par le CERN et soutenue par la Commission européenne. Elle permet aux chercheurs de déposer, conserver, diffuser et citer leurs travaux scientifiques en accès libre.

### **Objectifs de la plateforme Zenodo**

Les principaux objectifs de Zenodo sont :

- Favoriser la science ouverte en rendant les résultats de la recherche accessibles à tous.
- Assurer la conservation à long terme des données et publications scientifiques.
- Permettre la citation des travaux grâce à l'attribution d'un DOI (Digital Object Identifier).
- Valoriser les travaux de recherche, même ceux qui ne sont pas publiés dans des revues scientifiques.
- Faciliter le partage et la réutilisation des données entre chercheurs.

## Types de contenus hébergés sur Zenodo

Zenodo accepte une grande variété de contenus scientifiques, notamment :

- \* Articles scientifiques (préprints et postprints)
- \* Jeux de données (datasets)
- \* Thèses et mémoires
- \* Rapports techniques
- \* Présentations (PowerPoint, PDF)
- \* Logiciels et codes sources
- \* Images, vidéos et fichiers audio
- \* Posters scientifiques

Tous ces contenus peuvent être déposés gratuitement et en accès libre.

## Intérêt de Zenodo pour la science ouverte et la recherche en sciences de la nature et de la vie

- Facilite l'accès libre et gratuit aux données scientifiques (science ouverte).
- Encourage le partage et la réutilisation des données de recherche.
- Améliore la transparence et la reproductibilité des travaux scientifiques.
- Met à disposition des datasets en biologie (cellules, tissus, génomes).
- Soutient la formation des étudiants et la recherche scientifique.
- Attribue un DOI garantissant la traçabilité et la citation des données.
- Valorise le travail des chercheurs et des institutions.
- Il respecte les principes FAIR (Findable, Accessible, Interoperable, Reusable).

## 2. Description des étapes réalisées

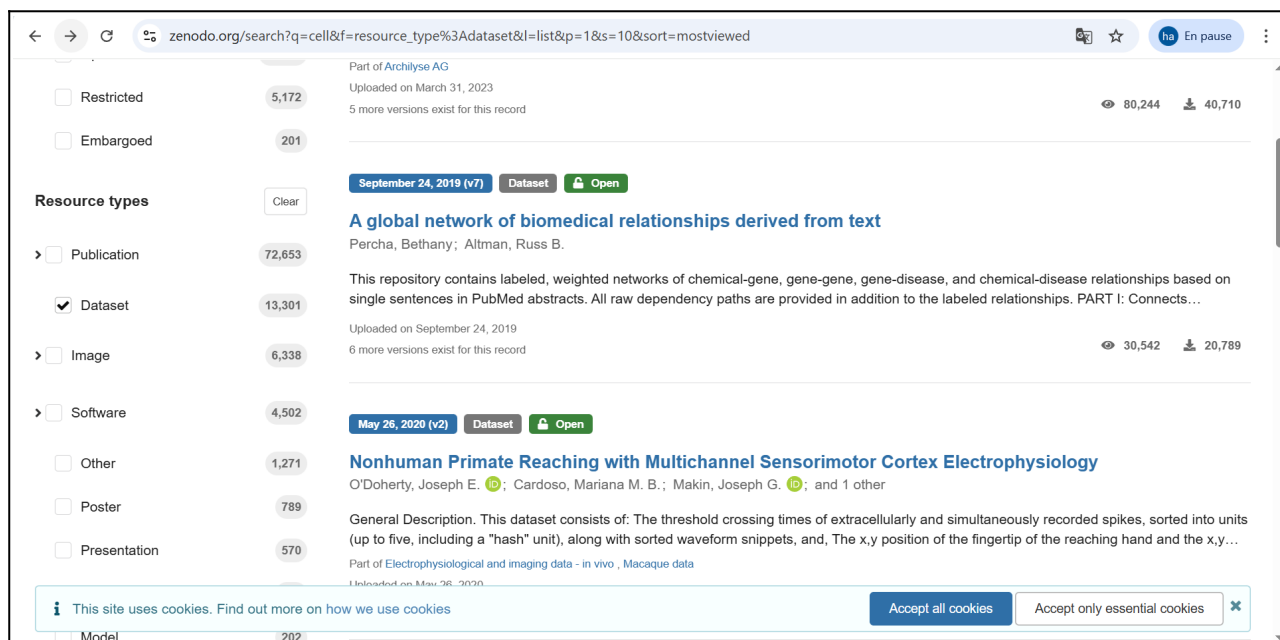
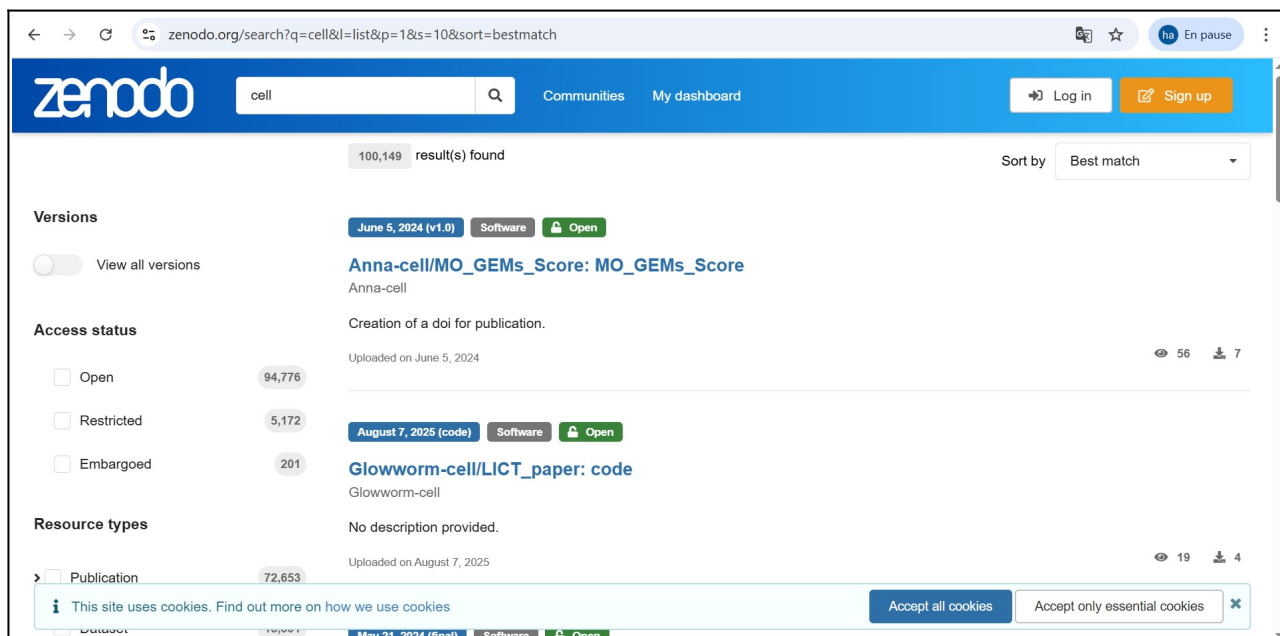
### Recherche effectuée (requête utilisée)

La plateforme Zenodo a été consultée via l'adresse officielle [<https://zenodo.org>].

Une recherche a été effectuée à l'aide de la barre de recherche principale en utilisant le mot-clé « cell », conformément aux consignes.

Cette requête a permis d'obtenir plusieurs jeux de données liés aux cellules.

Afin d'affiner les résultats, le type de document Dataset a été privilégié.



**Figure 1:** Page de recherche Zenodo montrant la requête « cell » et la liste des résultats obtenus.

## Critères de sélection du dataset

- Pertinence pour les sciences de la nature et de la vie
- Présence de métadonnées complètes
- Existence d'un DOI
- Données accessibles en libre accès
- Format exploitable pour une analyse scientifique

The screenshot shows the main page of a dataset on Zenodo. The title is "Training data for de novo transcriptome reconstruction from RNA-seq data". The authors listed are Heydarian, Mohammad<sup>1</sup>, Freeberg, Mallory<sup>1</sup>, and Poterlowicz, Krzysztof<sup>2</sup>. The publication date is May 24, 2017, and it is version v1. The page shows 15K views and 15K downloads. A description of the data is provided, mentioning it's part of a Galaxy Training Network tutorial. On the right, there are sections for Versions (showing v1), External resources (indexed in OpenAIRE), and a cookie consent banner at the bottom.

**Figure 2 :** Page principale du dataset sélectionné (titre, auteurs, date visibles).

## Navigation sur la plateforme

La page du dataset sélectionné permet d'accéder à :

- \* Une description détaillée du contenu
- \* Les fichiers associés
- \* Les métadonnées normalisées
- \* Les informations de licence

Le dataset a été téléchargé à l'aide du bouton Download, permettant l'accès aux fichiers de données.

This screenshot shows the 'Files' section of the dataset page. It contains a table with columns for 'Name', 'Size', and a 'Download' button for each file. The files are organized into pairs of forward and reverse reads for G1E and Megakaryocyte cell lines. The right sidebar shows the 'Communities' (Galaxy Training Network), 'Keywords and subjects' (Galaxy Training Network, useGalaxy, RNA-seq, hematopoiesis, Megakaryocytes, Mouse, Gata1-erythroid progenitor), and 'Details' (DOI: 10.123/GTndenovoRNAseq, Resource type: Dataset, Publisher).

Name	Size	Download
G1E_rep1_forward_read_(SRR549355_1)	277.3 MB	Download
G1E_rep1_reverse_read_(SRR549355_2)	277.3 MB	Download
G1E_rep2_forward_read_(SRR549356_1)	362.2 MB	Download
G1E_rep2_reverse_read_(SRR549356_2)	362.2 MB	Download
Megakaryocyte_rep1_forward_read_(SRR549357_1)	213.4 MB	Download
Megakaryocyte_rep1_reverse_read_(SRR549357_2)	213.4 MB	Download
Megakaryocyte_rep2_forward_read_(SRR549358_1)	12.7 MB	Download

**Figure 3 :** Section de téléchargement des fichiers du dataset.



### 3. Métadonnées du dataset

#### Norme de métadonnées utilisée

Les métadonnées du dataset ont été récupérées et organisées selon la norme Dublin Core, une norme internationale largement utilisée pour la description des ressources numériques scientifiques.

Cette norme facilite l'identification, la citation et la réutilisation des jeux de données.

Élément Dublin Core	Informations du dataset
<b>Titre</b>	Training data for de novo transcriptome reconstruction from RNA-seq data
<b>Auteurs</b>	1. <a href="#">Heydarian, Mohammad</a> 2. <a href="#">Freeberg, Mallor</a> 3.Poterlowicz, Krzysztof
<b>Sujet</b>	RNA-seq, transcriptome, reconstruction de novo, expression génique, cellules, hématopoïèse
<b>Description</b>	Données d'entraînement utilisées dans un tutoriel du Galaxy Training Network pour l'analyse RNA-seq par reconstruction de novo du transcriptome. Les données proviennent d'une étude de Wu et al. (2014) portant sur la régulation génique par le facteur de transcription Tal1 au cours de la différenciation hématopoïétique chez la souris.
<b>Éditeur (Publisher)</b>	Zenodo
<b>Contributeur (Contributor)</b>	Galaxy Training Network
<b>Date</b>	2017
<b>Type</b>	Dataset
<b>Format</b>	Fichiers RNA-seq (FASTQ, données downsampled)
<b>Identifiant unique (DOI)</b>	<a href="https://doi.org/10.123/GTNdenovoRNAseq">https://doi.org/10.123/GTNdenovoRNAseq</a>
<b>Source</b>	Étude de Wu et al., 2014 – GEO Accession: GSE51338
<b>Langue du contenu</b>	Anglais
<b>Couverture</b>	Cellules de souris (G1E et mégacaryocytes), différenciation hématopoïétique
<b>Droits d'utilisation</b>	Creative Commons Attribution 4.0 International



## **Références**

Biopython official website — <https://biopython.org/>

Biopython documentation — <https://biopython.org/wiki/Documentation>

Biopython Tutorial and Cookbook — <https://biopython.org/DIST/docs/tutorial/Tutorial.html>

Cock et al., “Biopython: freely available Python tools for computational molecular biology and bioinformatics”, *Bioinformatics*, 2009

NCBI – National Center for Biotechnology Information