



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



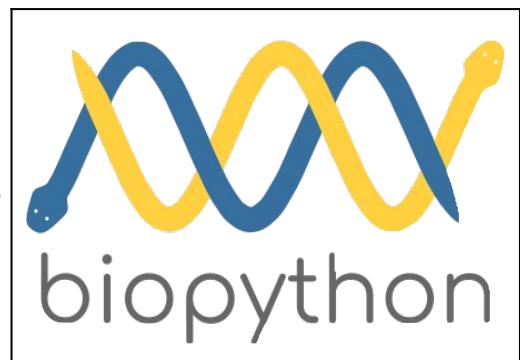
Module : Logiciels Libres et Open Source

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

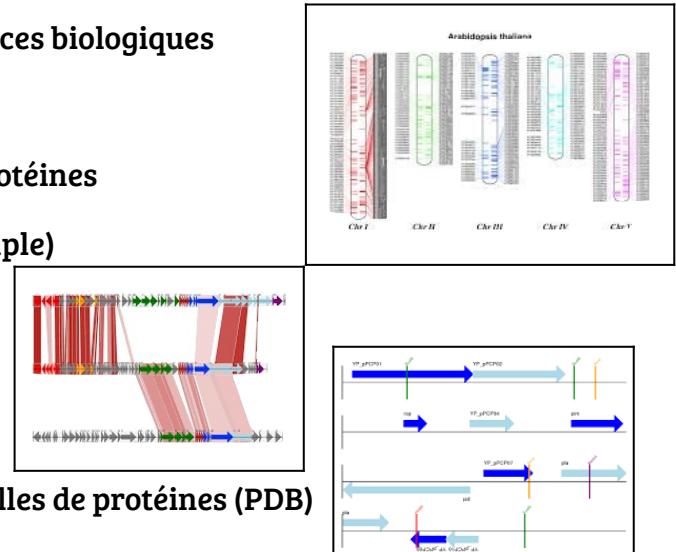


Figure 2 : Exemples de manipulation de séquences biologiques à l'aide de Biopython.

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**)

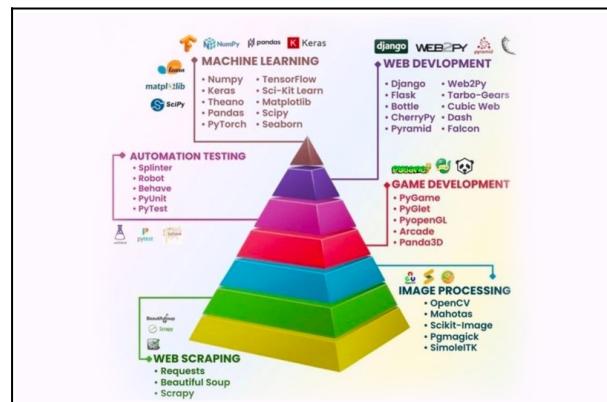


Figure 3: Architecture générale et intégration de Biopython avec l'écosystème Python.

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 structurelle.

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.

The screenshot shows the Zenodo search interface with the query "cell". The results are sorted by "Best match".

- Versions:** June 5, 2024 (v1.0) Software Open
- Anna-cell/MO_GEMs_Score: MO_GEMs_Score** (Anna-cell)
 - Creation of a doi for publication.
 - Uploaded on June 5, 2024
 - 56 views, 7 downloads
- Access status:** Open (94,776), Restricted (5,172), Embargoed (201)
- Resource types:** Publication (72,653), Dataset (1999), Software (May 21, 2024 (final))
- Details for a dataset:** Glowworm-cell/LICT_paper: code (Glowworm-cell)
 - No description provided.
 - Uploaded on August 7, 2025
 - 19 views, 4 downloads

At the bottom, there are buttons for "Accept all cookies" and "Accept only essential cookies".

The screenshot shows the Zenodo search interface with the query "cell" and a filter applied to "resource_type:Dataset". The results are sorted by "mostviewed".

- Access status:** Restricted (5,172), Embargoed (201)
- Resource types:** Publication (72,653), Dataset (13,301), Image (6,338), Software (4,502), Other (1,271), Poster (789), Presentation (570)
- Details for a dataset:** A global network of biomedical relationships derived from text (Percha, Bethany; Altman, Russ B.)
 - Part of Archilyse AG
 - Uploaded on March 31, 2023
 - 5 more versions exist for this record
 - 80,244 views, 40,710 downloads
- Details for another dataset:** Nonhuman Primate Reaching with Multichannel Sensorimotor Cortex Electrophysiology (O'Doherty, Joseph E.; Cardoso, Mariana M. B.; Makin, Joseph G.; and 1 other)
 - September 24, 2019 (v7) Dataset Open
 - This repository contains labeled, weighted networks of chemical-gene, gene-gene, gene-disease, and chemical-disease relationships based on single sentences in PubMed abstracts. All raw dependency paths are provided in addition to the labeled relationships. PART I: Connects...
 - Uploaded on September 24, 2019
 - 6 more versions exist for this record
 - 30,542 views, 20,789 downloads

At the bottom, there are buttons for "Accept all cookies" and "Accept only essential cookies".

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 a Zenodo dataset page. At the top, there's a header with a back arrow, forward arrow, refresh button, and a URL bar showing 'zenodo.org/records/583140'. Below the header is a logo for 'Galaxy Training Network'. The main title is 'Training data for de novo transcriptome reconstruction from RNA-seq data'. It was published on May 24, 2017, and is version v1. There are two download buttons: 'Dataset' and 'Open'. To the right, there are statistics: '15K VIEWS' and '15K DOWNLOADS', with a link to 'Show more details'. Below these are sections for 'Versions' (listing 'Version v1' with a date of 'May 24, 2017' and a DOI of '10.123/GTNdenovoRNAsed') and 'External resources' (listing 'Indexed in' and 'OpenAIRE'). A cookie consent banner at the bottom says 'This site uses cookies. Find out more on how we use cookies' with buttons for 'Accept all cookies' and 'Accept only essential cookies'.

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.

The screenshot shows the 'Files' section of the Zenodo dataset page. It lists several files:

- G1E_rep1_forward_read_(SRR549355_1) - md5:712324a599eb08c509f63e1879db0 - 277.3 MB - Download
- G1E_rep1_reverse_read_(SRR549355_2) - md5:8c71b47a480f8dad38b64861d1f1653ec - 277.3 MB - Download
- G1E_rep2_forward_read_(SRR549356_1) - md5:6cd4c921bae906289a2c2741d549a2 - 362.2 MB - Download
- G1E_rep2_reverse_read_(SRR549356_2) - md5:b504e15f427a7e8dd35b46842b44196 - 362.2 MB - Download
- Megakaryocyte_rep1_forward_read_(SRR549357_1) - md5:535b73ef41bf7412d8439519c671d - 213.4 MB - Download
- Megakaryocyte_rep1_reverse_read_(SRR549357_2) - md5:e48cc6b9582804870602a5689bfa2389 - 213.4 MB - Download
- Megakaryocyte_rep2_forward_read_(SRR549358_1) - md5:888f80ec68222e4789148b06cce83ec1 - 12.7 MB - Download

 To the right of the file list are sections for 'Communities' (listing 'Galaxy Training Network'), 'Keywords and subjects' (listing 'Galaxy Training Network', 'useGalaxy', 'RNA-seq', 'hematopoiesis', 'Megakaryocytes', 'Mouse', 'Gata1-erythroid progenitor'), and 'Details' (listing 'DOI' as '10.123/GTNdenovoRNAsed', 'Resource type' as 'Dataset', and 'Publisher' information).

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
Titre	Training data for de novo transcriptome reconstruction from RNA-seq data
Auteurs	1. Heydarian, Mohammad 2. Freeberg, Mallor 3.Poterlowicz, Krzysztof
Sujet	RNA-seq, transcriptome, reconstruction de novo, expression génique, cellules, hématopoïèse
Description	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.
Éditeur (Publisher)	Zenodo
Contributeur (Contributor)	Galaxy Training Network
Date	2017
Type	Dataset
Format	Fichiers RNA-seq (FASTQ, données downsampled)
Identifiant unique (DOI)	https://doi.org/10.123/GTNdenovoRNaseq
Source	Étude de Wu et al., 2014 – GEO Accession: GSE51338
Langue du contenu	Anglais
Couverture	Cellules de souris (G1E et mégacaryocytes), différenciation hématopoïétique
Droits d'utilisation	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