



Ministère de l'enseignement supérieur et la recherche scientifique



Université Mohammed Seddik ben Yahia- Jijel

- **Faculté**: des sciences de la nature et de la vie.
- **Département** : de microbiologie appliquée et sciences alimentaires.
- **Niveau** : 1^{er} année master microbiologie appliquée.
- **Module** : Logiciels libres et open source.

TP : Logiciels Libres et Open Source

Réalisé par :

- Bourouina Hana
- Boussadia Rokia
- Haddada Nadjat

L'année universitaire : 2025/2026



SOMMAIRE

Introduction :	p 05
L'objectif :	p 06
Partie I : L'étude théorique d'un outil	
1. Généralités sur Biopython	p 07
2. Les fonctionnalités	p 07
3. L'aspect technique	p 08
4. Les points forts	p 09
5. Les points faibles et les limites	p 09
6. Conclusion	p 09
Partie II : L'étude pratique d'un outil zenodo	
1. La recherche d'un dataset	p 11
2. Présentation de zenodo	p 13
3. Description des étapes réalisées	p 14
Partie III : GitHub	
Conclusion générale	p 17
Références bibliographiques	p 18



Liste des figures

Figure 01	Biopython.	P 07
Figure 02	Page d'accueil de zenodo.	P 11
Figure 03	Représentation d'une recherche de dataset.	P 11
Figure 04	Filtration pour la recherche d'un dataset.	P 11
Figure 05	Représentation de sélection d'un dataset.	P 12
Figure 06	Représentation de téléchargement d'un dataset.	P 12
Figure 07	Représentation des métadonnées par Dublin core.	P 12
Figure 08	Type de contenu hébergé sur zenodo.	P 14



Liste des tableaux

Tableau 01	Représentation des modules de Biopython et leurs fonctions.	P 08
Tableau02	Représentation des métadonnées associées au dataset à l'aide de Dublin core.	P 13

Introduction :

L'évolution rapide des sciences de la vie et de la nature au cours des dernières décennies est étroitement liée au développement des outils informatiques. L'augmentation massive des données biologiques issues des techniques modernes telles que le séquençage à haut débit, la génomique, la transcriptomique, la protéomique et la métabolomique a rendu indispensable l'utilisation de solutions informatiques performantes permettant le stockage, le traitement et l'analyse de grandes quantités de données biologiques de manière efficace et reproductible. (Ouzounis & Valencia, 2003 ; Luscombe et al., 2001)

Dans ce contexte, les logiciels informatiques occupent une place centrale dans la recherche en sciences de la vie et de la nature, notamment en bio-informatique, biologie moléculaire, écologie et biotechnologie. Les logiciels libres et open source présentent des avantages considérables, tels que la transparence du code, la possibilité de modification, l'accès gratuit et le développement collaboratif, favorisant ainsi l'innovation scientifique et le partage des connaissances. (Open Source Initiative, 2024)

Parmi les outils informatiques les plus utilisés en bio-informatique figure Biopython, une bibliothèque open source développée en langage Python. Elle offre un ensemble riche de modules dédiés à l'analyse des séquences biologiques, à la manipulation de fichiers standards (FASTA, GenBank), à l'accès aux bases de données biologiques et à l'automatisation des analyses bio-informatiques. Grâce à sa simplicité, sa flexibilité et sa compatibilité avec d'autres bibliothèques scientifiques, Biopython est largement adopté par la communauté scientifique. (Cock et al., 2009 ; Biopython Documentation, 2024)

Parallèlement à l'analyse des données, la gestion et le partage des résultats scientifiques sont devenus des éléments essentiels de la recherche moderne. Zenodo est une plateforme de science ouverte permettant le dépôt, le partage et la citation de jeux de données, de publications et de logiciels. Elle contribue à la diffusion du savoir scientifique et garantit la pérennité des données grâce à l'attribution de DOI, favorisant ainsi la reproductibilité et la visibilité des travaux de recherche. (Zenodo Documentation, 2024)

En complément, GitHub est une plateforme collaborative largement utilisée pour l'hébergement et le développement de codes sources open source. Elle permet le travail collaboratif, la gestion des versions et le suivi des modifications des projets scientifiques, en particulier dans le domaine de la bio-informatique. L'utilisation conjointe de GitHub et de Zenodo améliore la traçabilité, la reproductibilité et la valorisation des outils et des données scientifiques. (GitHub Documentation, 2024)

Ainsi, l'intégration des outils informatiques tels que Biopython, associée aux plateformes ouvertes Zenodo et GitHub, constitue aujourd'hui un pilier fondamental de la recherche en sciences de la vie et de la nature. Ce travail vise à illustrer l'importance des logiciels libres et open source dans l'analyse, la gestion et la diffusion des données scientifiques. (Cock et al., 2009 ; Open Source Initiative, 2024)

Objectifs de ce travail :

Ce travail vise à

- ✓ Découvrir et analyser les outils utilisés dans les sciences de la vie et de la nature.
- ✓ Comprendre les avantages et les limites de l'utilisation des outils libres.
- ✓ Apprendre à élaborer un rapport scientifique structuré, clair et appuyé par des illustrations.

Partie I : L'étude théorique de Biopython

1. Généralité sur Biopython :

Biopython est une initiative internationale regroupant des développeurs qui créent des outils Python librement accessibles pour la biologie moléculaire computationnelle. Python est un langage orienté objet, interprété et flexible, de plus en plus populaire pour le calcul scientifique grâce à sa syntaxe claire et sa facilité d'apprentissage. Il peut également être étendu avec des modules écrits en C, C++ ou FORTRAN.

Biopython fournit une ressource en ligne pour accéder aux modules, scripts et liens utiles pour les développeurs en bio-informatique.

L'objectif principale de Biopython est de faciliter l'utilisation de langage Python en bio-informatique en proposant des modules et classes réutilisables de haute qualité.
(Chang, J., et al ; 2020)



Figure 01 : Biopython. (Chang, J., et al ; 2020)

2. Les fonctionnalités principales :

Parmi ses fonctionnalités, Biopython propose :

- La lecture et l'écriture des différents formats de fichiers bioinformatiques (BLAST, Clustalw, FASTA, Genbank, etc.)
- L'accès à des bases de données en ligne comme celles du NCBI (BLAST, PUBMED) et Expasy, y compris SWISS-PROT.
- Des interfaces vers des programmes courants ou spécialisés (Clustalw, DSSP, MSMS, etc.)
- Manipulation des séquences biologiques (ADN, ARN, protéine) tels que la transcription et la traduction.
- L'analyse et la comparaison des séquences, y compris l'alignement des séquences.
- Analyse phylogénétique et structurale (structure tridimensionnelle des protéines)
- Une documentation complète pour faciliter le développement et l'apprentissage.
(Chang, J., et al ; 2020)

3. Aspect technique de Biopython :

Biopython est écrit en Python et conçu de manière à ce que chaque fonction soit dans un module indépendant, ce qui facilite son organisation et son utilisation. Il contient également des analyseurs prêts à l'emploi pour gérer différents formats de données biologiques tels que FASTA, GenBank et Clustalw, ce qui permet à l'utilisateur de lire et d'écrire facilement des données sans avoir besoin d'une programmation complexe. (**Chang, J., et al ; 2020**). Biopython regroupe plusieurs modules et outils permettant la manipulation et l'analyse des données biologiques, le tableau suivant représente les modules et les outils de Biopython et leurs fonctions :

Tableau01 : Représentation des modules de Biopython et leurs fonctions.

Module	Fonction
Bio.Seq	Représentation et manipulation des séquences biologiques.
Bio.SeqRecord	Stockage des séquences avec ses informations.
Bio.SeqIO	Lecture et écriture des fichiers de séquences biologiques et conversion entre formats.
Bio.AlignIO	Lecture et écriture des fichiers d'alignement de séquences.
Bio.Align	Création et analyse de séquences et extraction des informations comparatives.
Bio.Blast	Envoi des requêtes BLAST et analyse des résultats.
Bio.Entrez	Connexion aux bases de données NCBI.
Bio.PDB	Lecture et analyse des structures 3D des protéines à partir des fichiers PDB.
Bio.Phylo	Lecture et analyse et visualisation des arbres phylogénétiques.
Bio.motifs	Analyse des motifs et régions conservées dans l'ADN et les protéines.
Bio.Restriction	Etude des enzymes de restriction et localisation de leurs sites de coupe dans l'ADN.
Bio.PopGen	Analyse des données de génétique des populations.
Bio.Cluster	Analyse par regroupement des données biologiques.
Bio.Phenotype	Analyse des données sur les phénotypes des organismes.
Bio.KEGG	Accès aux données KEGG sur les voies métaboliques et réseaux biologiques.
Bio.ExPASy	Connexion aux services ExPASy pour analyser les informations sur les protéines et les enzymes.
Bio.Feature	Représentation des caractéristiques et localisation des éléments dans les séquences.

(Chang, J., et al ; 2020)

4. Points forts :

- Open source, gratuit, facile à apprendre et manipuler grâce à la simplicité du langage python. (**Chapman B., Chang J., 2000**)
- Capacité à lire et à écrire de nombreux formats de données biologiques différents.
- Communication avec des bases de données biologiques en ligne.
- Outils prêts à l'emploi pour l'analyse des données.
- Gain de temps grâce à de nombreuses fonctions prêtes à l'emploi.
- Utile pour les débutants.
- Modules prêts à l'emploi qui réduisent le temps de programmation.

(Chang, J., et al ; 2020)

5. Limites et points faibles :

- Certaines fonctions sont complexes et difficiles à comprendre pour les débutants. (**Number Analytics, 2025**)
- Performances limitées avec des ensembles de données volumineux. (**Number Analytics, 2025**)
- Dépendance à des outils externes tels que BLAST. (**Chang, J., et al ; 2020**)
- Prise en charge limitée de certains formats de fichiers. (**Number Analytics, 2025**)

6. Conclusion :

En conclusion, Biopython constitue un outil incontournable dans le domaine de la bioinformatique et des sciences de la vie et de la nature. En tant que bibliothèque libre et open source basée sur le langage Python, elle offre une large gamme de fonctionnalités permettant l'analyse, la manipulation et l'exploitation efficace des données biologiques, notamment les séquences d'ADN, d'ARN et de protéines. Sa simplicité d'utilisation, combinée à la richesse de ses modules et à sa compatibilité avec d'autres bibliothèques scientifiques, en fait un outil particulièrement adapté aussi bien aux chercheurs qu'aux étudiants.

Par ailleurs, l'aspect open source de Biopython favorise la transparence des méthodes, la reproductibilité des analyses et la collaboration scientifique à l'échelle internationale. Toutefois, malgré ses nombreux avantages, son utilisation peut nécessiter des compétences de base en programmation, ce qui peut représenter une limite pour certains utilisateurs non initiés. Néanmoins, la disponibilité d'une

documentation riche et d'une communauté active permet de surmonter ces difficultés et d'exploiter pleinement le potentiel de cet outil.

Ainsi, Biopython s'impose comme une solution fiable, flexible et évolutive pour le traitement des données biologiques, contribuant de manière significative à l'avancement de la recherche scientifique et à l'intégration des outils informatiques dans les sciences de la vie et de la nature.

Partie II : Exploration de zenodo

1. Recherche d'un dataset :

1.1. Visitation de plateforme zenodo (<https://zenodo.org/>) :

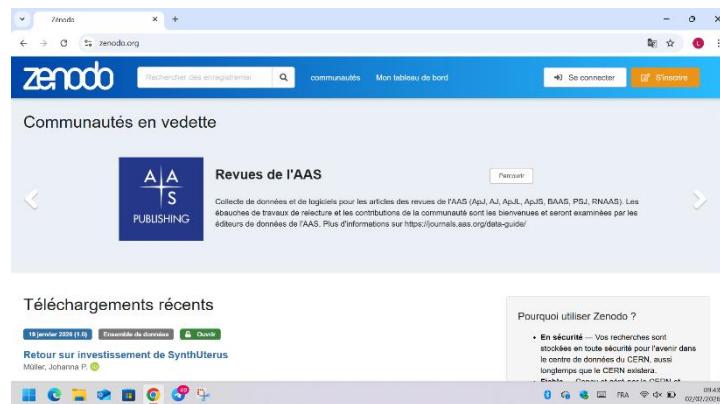


Figure 02 : page d'accueil de zenodo.

1.2. Une recherche de dataset à l'aide d'une requête contenant le mot clés : *Escherichia coli*

La barre de recherche

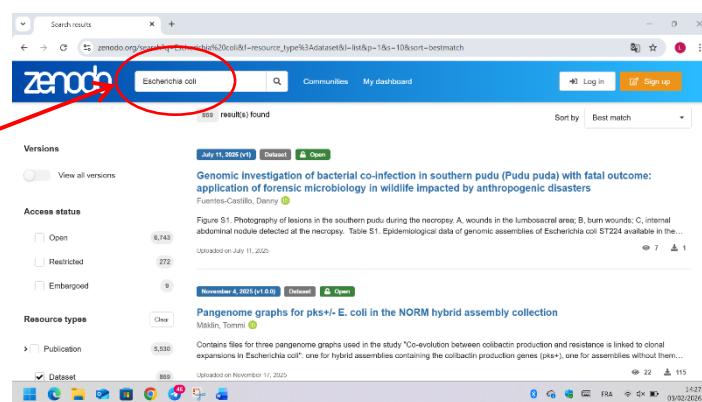


Figure 03 : représentation d'une recherche de dataset.

1.3. Une filtration de recherche a été effectuée :

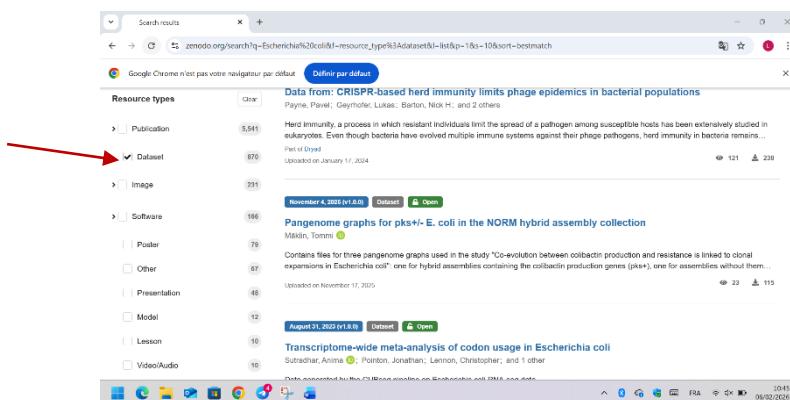


Figure 04 : filtration pour recherche d'un dataset.

1.4. La sélection d'un dataset pertinent :

The screenshot shows the Zenodo interface for the dataset "Escherichia coli OP50". The dataset was published on April 10, 2024, and has version 1.1. It has 79 views and 22 downloads. Contributors include Assele, Adrien, Blackburn, Diana, Zhang, Fan, and Supervisor, Samuel, Buck. The dataset is part of the WormBiome collection. A red arrow points to the "View all 2 versions" link in the "Versions" section.

Figure 05 : représentation de la sélection d'un dataset.

1.5. Le téléchargement de dataset choisi :

The screenshot shows the Zenodo interface for the dataset "Escherichia coli OP50". The "Files" section shows a single file "OP50.tar.bz2" (59.6 MB). A red circle highlights the "Download all" button. A red arrow points to the "External resources" section, which includes links for "Indri" and "OpenAIRE". A red line labeled "Cliquez ici" (Click here) points to the "Indri" link. The "Additional details" section provides metadata about the dataset, including subjects like Bacteria, Microbial genetics, Microbiome, Cenorhabditis elegans, and Host-bacteria relationships.

Figure 06 : représentation de téléchargement d'un dataset choisi

1.6. La récupération des métadonnées associées au dataset par Dublin core :

Dans l'option Export, sélectionnez **Dublin core XML**, puis exportez pour récupérer les métadonnées associées au jeu de données.

The screenshot shows the Zenodo interface for the dataset "Escherichia coli OP50" with the XML file "1095781.xml" open. The XML code represents the dataset's metadata in Dublin Core XML format. The XML structure includes elements like `<dc:identifier>`, `<dc:title>`, `<dc:subject>`, `<dc:type>`, `<dc:language>`, `<dc:publisher>`, `<dc:relation>`, `<dc:rights>`, `<dc:source>`, and `<dc:format>`.

Figure 07 : représentation des métadonnées par Dublin core.

1.7. Extraction des informations disponible :

Tableau 02 : représentation des métadonnées par Dublin core

Elément	Description
Titre	<i>Escherichia coli</i> OP50.
Créateur	Assie, Adrien.
Sujet	Bactérie, génome, base de données, wormbiome, génomique, annotations, nématode, hôte associée, <i>C.elegans</i> .
Description	Il s'agit d'un des fichiers d'archive de la base de données Wormbiome. Cette entrée comprend tous les fichiers d'annotation du génome relatifs à <i>Escherichia coli</i> OP50, une souche de Wormbiome.Gammaprotéobactéries. La collection Wormbiome est une base de données en ligne qui centralise toutes les informations relatives aux bactéries associées à <i>C. elegans</i> .
Editeur	Zenodo.
Contributeur	Dana Blackburn, Fan Zhang, S. Buck Samuel.
Date	10 Avril 2024.
Format	<u>OP50.tar.bz2</u> (59,6 MB).
Identifiant	DOI 10.5281/zenodo.10957881.
Source	Dataset.
Langue	Anglais.
Relation	Fait partie de : Collection de bases de données Wormbiome.
Couverture	Organisme étudié : <i>Escherichia coli</i> .
Droit	Accès libre- Licence Creative Commons Zero v 1.0 Universal.

2. Présentation de Zenodo :

2.1. L'objectif de la plateforme :

- Zenodo permet à tout chercheur de partager et préserver ses résultats scientifiques, incluant données, publications et logiciels. (**Plomp, 2025**). Zenodo est une plateforme de dépôt de données en libre accès, développée par le CERN. (**CERN, 2025**).
- Les fichiers peuvent être de tout type ou format, jusqu'à 50 Go, chaque dépôt recevant un DOI unique pour faciliter la citation. (**Plomp, 2025**).

- La plateforme soutient la science ouverte et héberge des contenus multidisciplinaires. (**CERN, 2025 ; Wikipedia contributors, 2025**).

2.2. Types de contenu hébergés sur Zenodo :

- Jeux de données (datasets) : des données brutes ou traitées issues de recherches scientifiques.
- Publications : des articles, rapports ou prépublications.
- Logiciels et codes sources : des programmes ou scripts utilisés dans la recherche.
- Présentations et posters : des contenus pour conférences ou enseignements.
- Images, vidéos et audio : des supports multimédias liés à la recherche.
- Autres artefacts numériques – tout matériel scientifique numérique ne rentrant pas dans les catégories ci-dessus.

(CERN, 2025 ; Plomp, 2025 ; Wikipedia contributors, 2025).

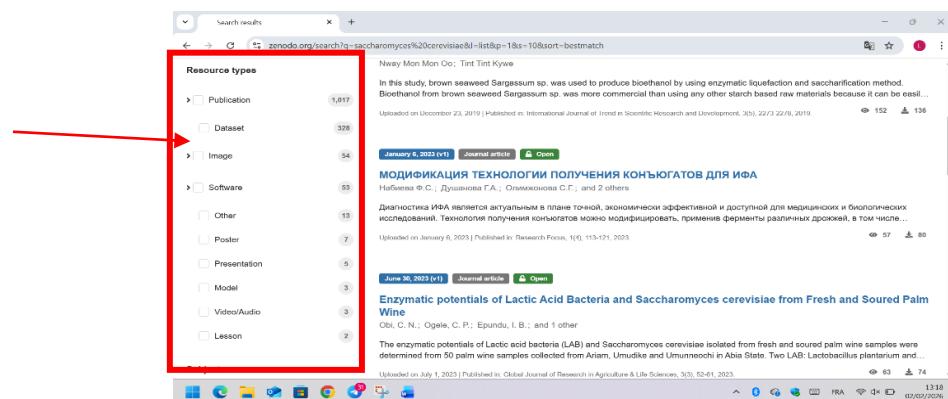


Figure 08 : type de contenu hébergé sur zenodo.

2.3. Intérêt de zenodo pour la science ouverte et la recherche en science de la nature et de la vie :

Zenodo joue un rôle important dans la diffusion de la science ouverte en favorisant la transparence, la reproductibilité des résultats scientifiques et la réutilisation des données particulièrement dans la recherche en science de la nature et de la vie. (**Pérez-Riverol et al., 2019**).

3. Description des étapes réalisées :

3.1. Recherche effectuée : La recherche est réalisée en utilisant la barre de recherche intégrée, des mots clés associées au thème et des filtres afin de faciliter la recherche. (**Plomp, 2025 ; CERN, 2025**)

3.2. Critères de sélection du dataset :

- Pertinence biologique.
- Accès libre et fiabilité de source.
- Open source, format et compatibilité. (**Plomp, 2025 ; CERN, 2025**)

3.3. Navigation sur plateforme :

- Page d'accueil : accès à la recherche, dépôt de fichiers et informations sur Zenodo.
- Recherche de contenu : barre de recherche pour trouver datasets, publications, logiciels, etc. Possibilité de filtrer par date, type, discipline ou licence.
- Dépôt de fichiers : bouton “Upload” pour télécharger vos données ou fichiers.
- L'utilisateur peut ajouter titre, description, auteurs, affiliations, licences et DOI.
- Téléchargement et citation : chaque dépôt possède un DOI unique, et les fichiers peuvent être téléchargés directement.

Partie III : GitHub

Conclusion générale :

Ce travail met en évidence l'importance des outils et plateformes numériques modernes dans le soutien de la science ouverte et le développement de la recherche scientifique en Sciences de la Nature et de la Vie. L'étude théorique de l'outil Biopython nous a permis de comprendre le rôle fondamental de la bio-informatique dans l'analyse et le traitement des données biologiques, notamment face à l'augmentation du volume et de la complexité des données scientifiques.

Par ailleurs, l'étude pratique de la plateforme Zenodo nous a permis de saisir l'importance de la mise à disposition des données scientifiques, de leur organisation à l'aide de standards de métadonnées, ainsi que de la garantie de leur réutilisation, contribuant ainsi à la transparence et à la collaboration scientifique.

Enfin, l'utilisation de GitHub nous a permis d'acquérir des compétences essentielles en matière de documentation, de partage et d'organisation des travaux scientifiques, en adéquation avec les exigences de la recherche scientifique moderne. Ce travail a ainsi contribué au développement de nos capacités méthodologiques et techniques, tout en renforçant le lien entre les aspects théoriques et pratiques de la science ouverte.



Références bibliographiques :

- 1. Biopython Documentation officielle.** <https://biopython.org>.
- 2. CERN.** (2025). *Zenodo – About the platform*. Retrieved from <https://about.zenodo.org>.
- 3. Chapman, B., & Chang, J.** (2000). *Biopython: Python tools for computational biology*. ACM SIGBIO Newsletter, 20(2), 15–19.
- 4. Cock, P. J. A., et al.** (2009). *Biopython: freely available Python tools for computational molecular biology and bioinformatics*. Bioinformatics, 25(11), 1422–1423.
- 5. GitHub Documentation.** <https://docs.github.com>.
- 6. Luscombe, N. M., et al.** (2001). *What is bioinformatics? A review*.
- 7. NumberAnalytics.** (2025). *Ultimate Guide to Biopython*. Retrieved from <https://www.numberanalytics.com/blog/ultimate-guide-biopython-introbioinformatics>.
- 8. Open Source Initiative.** <https://opensource.org>.
- 9. Ouzounis, C. A., & Valencia, A.** (2003). *Early bioinformatics: the birth of a discipline*. Bioinformatics.
- 10. Pérez-Riverol, Y., et al.** (2019). *The European Open Science Cloud: Supporting data-intensive research in Europe*. Nature Methods, 16(7), 595–597.
- 11. Plomp, E.** (2025). *Zenodo – Why, what and how?* Zenodo. <https://doi.org/10.5281/zenodo.14852437>.
- 12. Wikipedia contributors.** (2025). *Zenodo*. In Wikipedia, The Free Encyclopedia. Retrieved from <https://en.wikipedia.org/wiki/Zenodo>.
- 13. Zenodo Documentation.** <https://zenodo.org>.