



**Université Mohammed Seddik Ben Yahia
_Jijel**



**Faculté de SNV
Département de BMC
Spécialité : TFA
Module : Logiciels Libres et Open Source**

TP

Logiciels Libres et Open Source

Réalisé par :

Birem Meriem

Prof : Bensalem

Février 2026

PARTIE 1 – Étude théorique d'un outil

Présentation générale de BioPython



BioPython est une bibliothèque open-source écrite en langage Python, dédiée au traitement, à l'analyse et à l'exploitation des données biologiques en bio-informatique et en biologie computationnelle. Développée initialement par Chapman et Chang, BioPython a pour objectif principal de fournir des outils simples, robustes et réutilisables pour la manipulation des séquences biologiques, des structures moléculaires et des données issues des bases biologiques. Elle permet notamment la lecture, l'écriture et la conversion des formats standards (tels que FASTA, GenBank, EMBL, Swiss-Prot et PDB), ce qui facilite l'intégration des données hétérogènes issues de différentes plateformes expérimentales.

BioPython intègre également des modules puissants pour l'analyse de séquences (calcul de compositions, traduction ADN—protéine, recherche de motifs), l'alignement de séquences via des interfaces avec des outils reconnus comme BLAST et ClustalW, ainsi que l'analyse phylogénétique. En outre, le module Bio.Entrez permet un accès programmatique direct aux bases de données du NCBI, offrant ainsi la possibilité d'automatiser la récupération et l'analyse de grands volumes de données génomiques. Grâce à son intégration avec l'écosystème scientifique Python (NumPy, SciPy, Matplotlib), BioPython constitue un outil central pour le développement de pipelines bio-informatiques reproductibles et performants.

Fonctionnalités principales de Biopython

1. Manipulation et analyse des séquences biologiques

```
>gi|2765658|emb|Z78533.1|CIZ78533 C.irapeanum 5.8S rRNA gene and ITS1 and ITS2 DNA
CGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTGATGAGACCGTGGAATAAACGATCGAGTG
AATCCGGAGGACCGGTGTA CTACAGCTACCGGGGGCATTGCTCCCGTGGTGACCCGATTTGTTGTTGGG
CCGCCTCGGGAGCGTCCATGGCGGGTTTGAACCTCTAGCCCGGCGCAGTTTGGGCGCCAAGCCATATGAA
AGCATCACCGGCGAATGGCATTGTCTTCCCCAAAACCGGAGCGGCGGCGTGCTGTCGCGTGCCCAATGA
ATTTTGATGACTCTCGCAAACGGGAATCTTGGCTCTTTGCATCGGATGGAAGGACGCAGCGAAATGCGAT
AAGTGGTGTGAATTGCAAGATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAGGCCATCA
GGCTAAGGGCACGCCTGCTTGGGCGTCGCGCTTCGTCTCTCTCGCAATGCTTGCCCGGCATACAGCC
AGGCCGGCGTGGTGGGATGTGAAAGATTGGCCCTTGTGCCTAGGTGCGGCGGGTCCAAGAGCTGGTGT
TTTGATGGCCCGGAACCGGCAAGAGGTGGACGGATGCTGGCAGCAGCTGCCGTGCGAATCCCCATGTT
GTCGTGCTTGTGCGACAGGCAGGAGAACCCTTCCGAACCCCAATGGAGGGCGGTTGACCGCCATTGCGAT
GTGACCCAGGTGAGCGGGGGCACCCGCTGAGTTTACGC
```

Biopython excelle dans la manipulation de séquences biologiques (ADN, ARN, protéines) via des objets comme `bioSeq` et `SeqRecord` permettant d'effectuer des opérations fondamentales telles que la transcription, la traduction, le calcul de la composition en nucléotides ou acides aminés, ainsi que la recherche de motifs biologiques.

```
python
```

```
import Bio
from Bio import SeqIO
for seq_record in SeqIO.parse("ls_orchid.fasta", "fasta"):
    print(seq_record.id)
    print(repr(seq_record.seq))
    print(len(seq_record))
```

2. Lecture, écriture et conversion des formats bio-informatiques

En raison de module `Bio.SeqIO`, BioPython permet la lecture et l'écriture de nombreux formats standards (FASTA, GenBank, EMBL, Swiss-Prot, PDB). Cette capacité facilite l'interopérabilité entre différents logiciels bio-informatiques et garantit une manipulation fiable des données issues de sources multiples.

```
from Bio import SeqIO
record = SeqIO.read("genome.fasta", "fasta")
```

3. Alignement et comparaison de séquences

```
python
```

```
> from Bio import Align
> aligner = Align.PairwiseAligner()
> seq1 = "GAACT"
> seq2 = "GAT"
> score = aligner.score(seq1, seq2)
> score
3.0
> alignments = aligner.align(seq1, seq2)
> for alignment in alignments:
...     print(alignment)
GAACT
||--|
GA--T
<BLANKLINE>
GAACT
|-|-|
G-A-T
<BLANKLINE>
```

BioPython intègre des outils pour l'alignement pair-à-pair et multiple de séquences via le module Bio.Align, ainsi que des interfaces avec des logiciels externes largement utilisés tels que BLAST et ClustalW.

4. Accès aux bases de données biologiques

Le module Bio.Entrez fournit un accès programmatique direct aux bases de données du NCBI, permettant la recherche, le téléchargement et l'analyse automatisée de séquences et de métadonnées biologiques.

5. Analyse phylogénétique

Le module Bio.Phylo permet la lecture, l'écriture, la visualisation et la manipulation d'arbres phylogénétiques. Il supporte plusieurs formats standards (Newick, Nexus) et facilite l'interprétation des relations évolutives entre espèces ou séquences biologiques.

6. Analyse des structures moléculaires

Inclut le module Bio.PDB, dédié à la manipulation et à l'analyse des structures tridimensionnelles des protéines issues de la Protein Data Bank (PDB).

7. Intégration avec l'écosystème scientifique Python

BioPython est entièrement compatible avec les bibliothèques scientifiques Python telles que NumPy, SciPy et Matplotlib, ce qui permet l'analyse statistique avancée et la visualisation des résultats biologiques.

Aspects techniques de BioPython

Sur le plan technique, Biopython repose sur :

- Une architecture modulaire spécialisée (Seq, SeqIO, Align, Phylo, PDB, Entrez), ce qui facilite la maintenance, la réutilisation et l'extension du code ;
- Elle prend en charge les formats bio-informatiques standards (FASTA, GenBank, EMBL, FASTQ, PDB) grâce à des parseurs efficaces, capables de traiter de grands volumes de données avec une gestion optimisée de la mémoire ;
- Une compatibilité avec les principaux systèmes d'exploitation (Windows, Linux, macOS) ;
- BioPython propose des interfaces avec des outils externes majeurs (BLAST, ClustalW, MUSCLE), permettant l'automatisation des analyses et leur intégration dans des pipelines reproductibles. Le module Entrez assure un accès distant aux bases de données du NCBI, conforme aux contraintes techniques des services web ;
- La bibliothèque fournit des outils dédiés à l'analyse des alignements, des arbres phylogénétiques et des structures protéiques 3D, avec des structures de données adaptées à chaque type d'analyse ;

- Une intégration avec d'autres bibliothèques scientifiques Python telles que NumPy, SciPy et Matplotlib ;
- Une compatibilité avec les principaux systèmes d'exploitation (Windows, Linux, macOS) ;

Points forts

- ✓ Logiciel libre et open source et gratuit
- ✓ Facilité d'utilisation, grâce à la simplicité du langage Python ;
- ✓ Grande flexibilité, permettant de personnaliser les analyses selon les besoins de l'utilisateur ;
- ✓ Large communauté d'utilisateurs et de développeurs, garantissant un support continu ;
- ✓ Interopérabilité, avec de nombreux outils et bases de données bioinformatiques ;
- ✓ Reproductibilité scientifique, grâce à l'automatisation des analyses.

Limites et points faibles

Malgré ses nombreux avantages, Biopython présente certaines limites :

- Absence d'interface graphique, ce qui peut représenter un obstacle pour les utilisateurs non familiers avec la programmation ;
- Performances dépendantes de Python, moins rapides que certains langages compilés pour des analyses à très grande échelle ;
- Dépendance à des outils externes, pour certaines analyses avancées (BLAST, alignements multiples) ;
- Courbe d'apprentissage initiale, notamment pour les utilisateurs sans expérience en programmation.

Conclusion

BioPython est un outil puissant et flexible pour la bio-informatique, offrant des fonctionnalités complètes pour l'analyse des séquences, la phylogénie et l'accès aux bases de données. Facile à utiliser grâce à Python et son écosystème, il reste limité par ses performances pour les gros jeux de données et certaines analyses avancées. Globalement, il constitue un outil incontournable, à la fois accessible et efficace, pour la recherche scientifique.

Les références bibliographiques

- Bloch, L. (2019). Python et BioPython. Bloch (site personnel, ISSN 2271-3905)
- Chapman B., Chang J. (2000). Biopython: Python tools for computational biology. ACM SIGBIO Newsletter, 20(2), 15–19.
- Cock P.J.A., et al. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Research, 38(6), 1767–1771.
- Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., & de Hoon, M. J. L. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics.
- Dameron, O., & Farrant, G. (2014, juillet). La bioinformatique avec BioPython. GNU/Linux Magazine, HS n°73. GNU/Linux Magazine Connect
- Millman, K. J., Aivazis, M. (2011). Python for Scientists and Engineers. Computing in Science & Engineering, 13(2), 9–12.
- NCBI Resource Coordinators (2018). Database resources of the National Center for Biotechnology Information.
- Perkel, J. M. (2015). Programming: Pick up Python. Nature, 518, 125–126

PARTIE 2 – Étude pratique : exploration de Zenodo

1. Présentation de Zenodo

❖ Objectifs de la plateforme

Zenodo est une plateforme d'archivage numérique ouverte et gratuite, développée par le CERN (European Organization for Nuclear Research), issu d'une collaboration avec OpenAIRE. Son objectif principal est de permettre aux chercheurs de partager, préserver et citer facilement tous types de résultats de recherche, y compris les publications, données, logiciels et autres artefacts numériques, surtout quand aucun dépôt dédié n'existe

Zenodo soutient les politiques d'Open Access et d'Open Data de l'Union européenne en offrant un entrepôt multidisciplinaire accessible à tous les domaines scientifiques. Elle attribue un identifiant persistant (DOI) à chaque dépôt pour assurer la citabilité et la pérennité des contenus.

❖ Types de contenus hébergés

Zenodo héberge une grande variété de contenus de recherche, couvrant tous les types de résultats numériques. Cela inclut publications, données et logiciels, ainsi que supports multimédias et autres artefacts...

- Publications : Articles scientifiques, posters, présentations et rapports.
- Données : datasets (Ensembles de données) et métadonnées.
- Logiciels : Code source, scripts et exécutables.
- Multimédia : Images, vidéos, audio et fichiers 3D

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

Zenodo joue un rôle clé dans la science ouverte en favorisant l'accès libre et la réutilisation des résultats de recherche. Pour la recherche en sciences de la nature et de la vie (SVT), elle offre un archivage fiable des données biologiques, écologiques et environnementales souvent massives et multidisciplinaires.

Dans les domaines des sciences de la nature et de la vie, Zenodo :

- Facilite le partage de **données biologiques et génomiques**
- Améliore la **reproductibilité des résultats**
- Permet la **réutilisation des données** par d'autres chercheurs
- Soutient les principes **FAIR** (Findable, Accessible, Interoperable, Reusable)

2. Description des étapes réalisées

❖ Recherche effectuée (requête utilisé)

Une recherche a été effectuée sur la plateforme Zenodo à l'aide du mot-clé « cell ». Un dataset pertinent a été sélectionné selon des critères scientifiques et d'accessibilité.

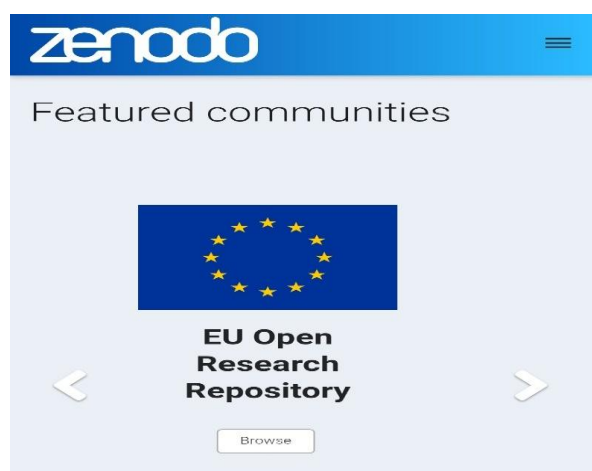


Figure 1

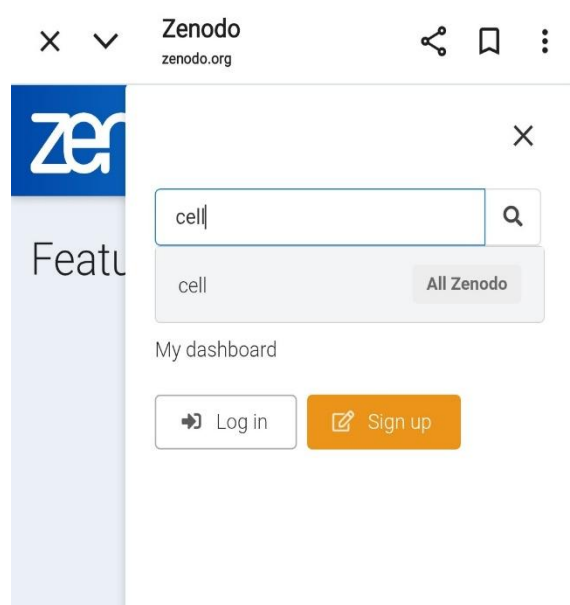


Figure 2

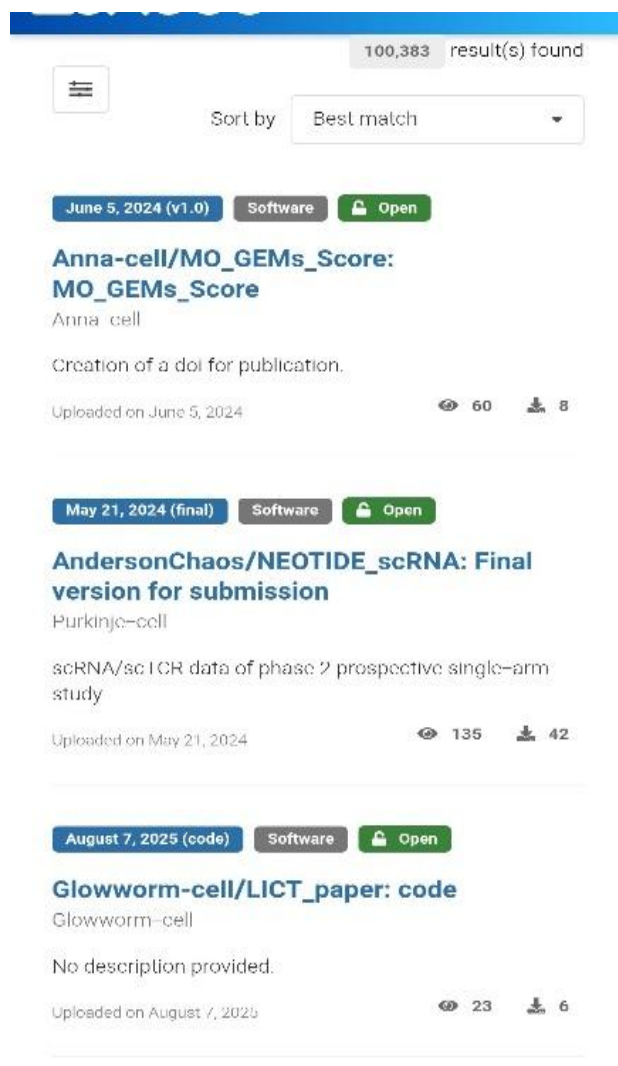


Figure 3

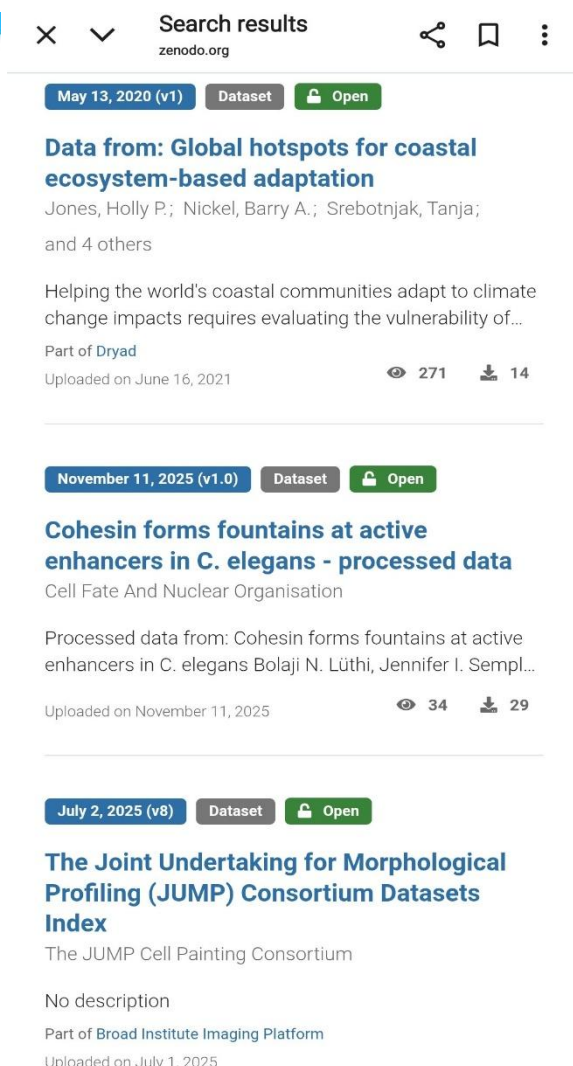




Figure 4

❖ Critères de sélection du dataset




Le dataset a été sélectionné selon les critères suivants :



- Pertinence pour les **sciences de la vie**
- Présence de **métadonnées complètes**
- Accès libre (open access)
- Utilisation claire d'une norme de métadonnées (Dublin Core)

❖ Dataset sélectionné




Cohesin forms fo...
zenodo.org






Published November 11,
2025 | Version v1.0

Dataset

 Open

Cohesin forms fountains at active enhancers in *C. elegans* – processed data

Cell Fate And Nuclear Organisation (Other)¹ 

Show affiliations

Processed data from:

Cohesin forms fountains at active enhancers in *C. elegans*
Bolaji N. Lüthi, Jennifer I. Semple, Anja Haemmerli, Saurabh Thapliyal, Kalyan Ghadage, Klement Stojanovski, Dario D'Asaro, Moushumi Das, Nick Gilbert, Dominique A. Glauser, Benjamin Towbin, Daniel Jost, Peter Meister *bioRxiv* 2023.07.14.549011;
doi: <https://doi.org/10.1101/2023.07.14.549011>

[now accepted in Nature Communications]

HiC data:

20221213_366_C_merge_2000.mcool – Control worms

20221213_366_A_merge_2000.mcool – Alpha amanitin treated worms

bTMP pulldown:

Figure 5

bTMP pulldown:

bTMP_WT_L3_subsampled_inputnormalized_minusgDNA_Znorm_mean.bw – Control worms

bTMP_Amanitintreated_L3_subsampled_inputnormalized_minusgDNA_Znorm_mean.bw – Alpha amanitin treated worms

Files









| Files (1.4 GB)  | |
|---|--|
| 20221213_366_A_merge_2000.mcool md5:0b57a2fdf1e0070754ab86bcdce49344  | |
| 328.8 MB | |
| |  Download |
| 20221213_366_C_merge_2000.mcool md5:f38ecef36c492fcab37ee1326e45362b  | |
| 301.6 MB | |
| |  Download |
| bTMP_Amanitintreated_L3_subsampled_inputnormalized_minusgDNA_Znorm_mean.bw md5:4026e4508cc5ea8cee5f419e61746d69  | |
| 392.8 MB | |
| |  Download |
| bTMP_WT_L3_subsampled_inputnormalized_minusgDNA_Znorm_mean.bw md5:f013827b61a6c96e2a28f79a371d8630  | |


Figure 6

Additional details

Related works

Is supplement to

Dataset: [10.1101/2023.07.14.549011](https://doi.org/10.1101/2023.07.14.549011) (DOI)

Citations 

▼

Show only:

↓

Search

☐ Literature (0)

☐ Dataset (0)


☐ Software (0)

☐ Unknown (0)


☐ Citations To This Version

No citations found

34

 VIEWS

29

 DOWNLOADS

► Show more details

Figure 7

Versions

Version v1.0

Nov 11, 2025

10.5281/zenodo.17582193

Cite all versions? You can cite all versions by using the DOI [10.5281/zenodo.17582192](https://doi.org/10.5281/zenodo.17582192). This DOI represents all versions, and will always resolve to the latest one. [Read more](#).

External resources

Indexed in



Details

DOI

DOI [10.5281/zenodo.17582193](https://doi.org/10.5281/zenodo.17582193)



Resource type

Dataset

Publisher

Zenodo

Rights

License



Creative Commons Attribution 4.0
International

Figure 8

Rights

License



Creative Commons Attribution 4.0
International

Citation

Cell Fate And Nuclear Organisation. (2025). Cohesin forms fountains at active enhancers in *C. elegans* – processed data (v1.0) [Data set]. Zenodo.

<https://doi.org/10.5281/zenodo.17582193>

Style

APA



Export

JSON



Export



Technical metadata

Created November 11, 2025

Modified November 11, 2025

Figure 9

❖ Téléchargement de dataset



Figure 10

Titre : Cohesin forms fountains at active enhancers in *C. elegans* – processed data

Type : Dataset

Date de publication : 11 novembre 2025

Accès : Open Access

3. Métadonnées du dataset

Les métadonnées ont été récupérées depuis la page du dépôt Zenodo et organisées selon la norme Dublin Core.

Tableau : Présentation structuré des métadonnées (Dublin core)

| | |
|------------|---|
| Title | Cohesin forms fountains at active enhancers in <i>C. elegans</i> - processed data |
| creator | Cell Fate And Nuclear Organisation |
| publisher | Zenodo |
| Date | 11 novembre 2025 |
| Type | Dataset |
| Format | ZIP/FASTA |
| Identifier | 10.5281/zenodo.17582193 |
| Right | Licence : Creative Commons Attribution 4.0 International |

Conclusion

Cette pratique a permis de faire la découverte de cette plateforme appelée Zenodo ainsi que son importance pour la communauté scientifique en termes de partage des données scientifiques. L'expérience sur le dataset génomique a permis d'envisager l'importance de la norme sur les métadonnées, comme Dublin Core, pour la traçabilité des sciences de vie. Zenodo représente aujourd'hui l'un des outils incontournables en matière de recherche moderne et de 'science ouverte

Les références bibliographiques

- CERN – Organisation européenne pour la recherche nucléaire. (2024).
Zenodo : plateforme de partage et d’archivage des données de recherche.
- Cell Fate and Nuclear Organisation. (2025).
Cohesin forms fountains at active enhancers in *C. elegans* – données traitées (Version 1.0) [Jeu de données]. Zenodo.
- Commission européenne. (2020). La science ouverte et le Cloud européen pour la science ouverte (EOSC). Office des publications de l’Union européenne.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016).
Les principes directeurs FAIR pour la gestion et la conservation des données scientifiques.
Scientific Data, 3, 160018.