

République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieur et de la recherche scientifique

University Mohammed Seddik ben Yahya _Jijel



Faculté : Sciences De La Nature Et De La Vie

Département : Biologie Moléculaire et Cellulaire

Matière : Logiciels et open source

Master 1 Biochimie

TP : Logiciels et open source

Préparer par :

- Meriem Himeur
- Bourouh Selssabil
- Zeghbib Oumeima

Année Universitaire 2025/2026

Plan du travail:

Partie 01:

Qu'est-ce que Biopython ?.....	04
Objectifs principaux.....	04
Fonctionnalités clés.....	04
Aspects techniques.....	05
Points forts	06
Limites et points faibles.....	06
Conclusion.....	07

Partie 02 :

Présentation de la plateforme Zenodo	09
Objectifs principaux de Zenodo.....	09
Types de contenus hébergés sur Zenodo.....	09
Intérêt de Zenodo pour la science ouverte et la recherche.....	10
Pourquoi Zenodo est utile pour les sciences de la nature et de la vie?.....	10
Description des étapes réalisées	10
Métadonnées du dataset	13

Etude théorique de Biopython



I. Qu'est-ce que Biopython ?:

Biopython est une **collection de modules Python open-source** conçue pour la **bioinformatique et la biologie computationnelle**.

Elle fournit des outils pour **manipuler, analyser et transformer des données biologiques** telles que les séquences d'ADN, d'ARN et de protéines. [1]

II. Objectifs principaux

- Lire, traiter et analyser des séquences biologiques (ADN, ARN, protéines). [1]
- Connecter et interagir avec des bases de données biologiques en ligne (comme NCBI). [1]
- Intégrer et automatiser des tâches bioinformatiques avec Python [1]

III. Fonctionnalités clés

1. Manipulation de séquences

Représente les séquences biologiques grâce à la classe Seq. Permet des opérations biologiques courantes :

- **transcription ADN → ARN**

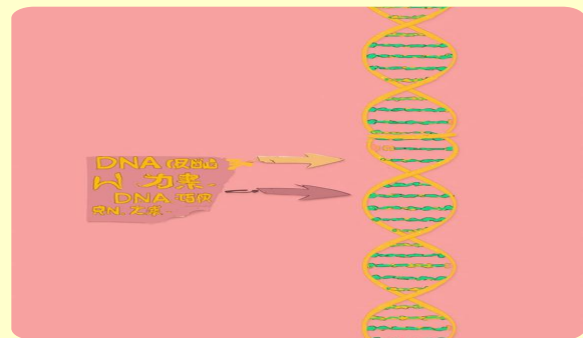


Figure 01 : Transcription ADN → ARN

- **traduction en protéines** [1]

2. Lecture et écriture de formats de fichiers biologiques

Biopython sait lire et écrire de nombreux formats utilisés en bioinformatique :

FASTA, GenBank, BLAST, Clustal, etc. Cela facilite l'importation et la conversion de données entre formats. [1]

3. Accès aux bases de données en ligne

Biopython intègre des modules pour interagir avec des services externes comme :

- les bases NCBI (Entrez, BLAST), [2]
- d'autres ressources biologiques en ligne. [2]

Cela permet de **télécharger des séquences et métadonnées automatiquement** dans vos scripts Python.

4. Outils avancés

Biopython inclut aussi des capacités pour :

- l'**alignement de séquences** . [3]
- l'**analyse phylogénétique** . [4]
- la **bioinformatique structurale** (parsing de structures de protéines). [5]

🔗 Exemple simple de manipulation de séquence :

Ce petit exemple illustre un usage courant de Biopython :

```
from Bio.Seq import Seq  
  
sequence = Seq("AGTACACTGGT")  
  
print("Séquence :", sequence)  
  
print("Transcription :", sequence.transcribe())  
  
print("Traduction :", sequence.translate())
```

IV. Aspects techniques

1. Architecture modulaire

Biopython est organisé en plusieurs *modules spécialisés* couvrant différents domaines de la bioinformatique (séquences, alignements, structures, phylogénie, etc.) qui peuvent être utilisés indépendamment ou ensemble dans un même script. [3]

2. Support de formats biologiques courants

La bibliothèque inclut des parseurs robustes capables de lire/écrire des formats très utilisés en bioinformatique tels que FASTA, FASTQ, GenBank, BLAST, PDB/mmCIF et autres. [3]

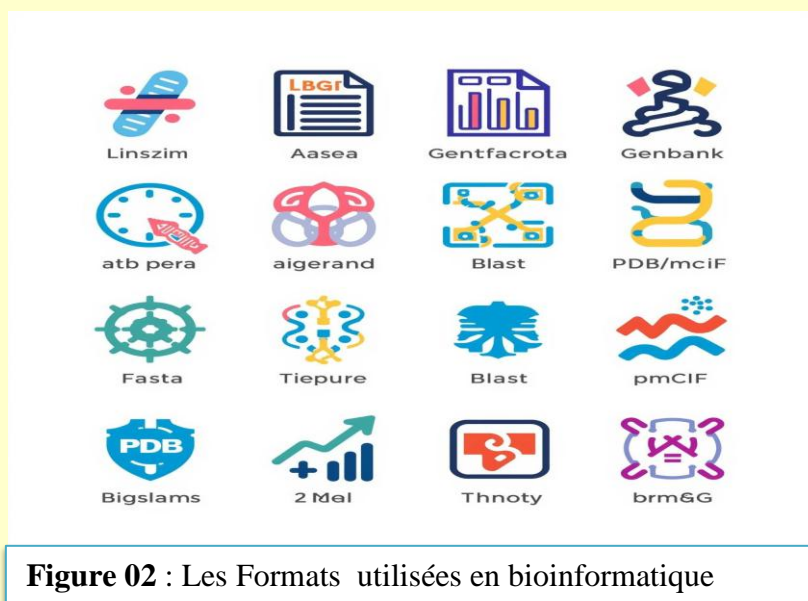


Figure 02 : Les Formats utilisées en bioinformatique

3. Manipulation et analyse de séquences

Elle fournit des structures de données biologiques comme `Seq` et `SeqRecord` permettant de manipuler des séquences, annotations et méta-données, ainsi que des outils pour les opérations biologiques de base. [3]

4. Intégration avec des outils externes

Biopython inclut des wrappers pour exécuter et analyser les résultats d'outils externes comme **BLAST**, **ClustalW**, **EMBOSS**, et pour interagir avec des services de bases de données tels que NCBI Entrez. [3]

5. Interopérabilité avec l'écosystème Python

La bibliothèque fonctionne de façon fluide avec d'autres outils scientifiques Python comme **NumPy**, **Matplotlib**, **SciPy**, ce qui facilite l'analyse et la visualisation de données. [3]

V. Points forts

1. Large couverture fonctionnelle

Biopython couvre une grande variété de tâches incontournables en bioinformatique — lecture de séquences, alignements, phylogénie, structures — ce qui en fait une bibliothèque *complète et polyvalente*. [3]



2. Facilité d'utilisation

Les interfaces sont conçues pour être simples à apprendre pour des utilisateurs Python, ce qui accélère la productivité des chercheurs et développeurs. [3]



3. Communauté active et maintenue

Le projet est open source, maintenu par une large communauté internationale, avec des mises à jour régulières, une documentation et des listes de diffusion. [3]



4. Grande compatibilité

Compatible avec Python 3 et avec des packages scientifiques standard, ce qui facilite son intégration dans des pipelines existants. [7]

VI. Limites et points faibles

1. Performance pour gros ensembles de données

Parce que Python est un langage interprété, certains traitements particulièrement lourds (par ex. alignements de très grandes séquences) peuvent être plus lents qu'avec des outils spécialisés en C/C++. [8]

2. Support de formats spécialisés

Biopython supporte la plupart des formats standard mais peut ne pas couvrir des formats propriétaires ou très spécifiques pour certaines tâches avancées. [8]

3. Courbe d'apprentissage

Certaines parties de l'API peuvent être difficiles pour les débutants, en particulier les modules avancés ou l'usage de wrappers d'outils externes qui nécessitent compréhension des paramètres bioinformatiques. [9]

4. Qualité et organisation de documentation

La documentation n'est pas toujours perçue comme facile à parcourir pour certaines sections ou modules plus obscurs du package. [9]

VII. Conclusion

Biopython est **un outil central et incontournable** en bioinformatique pour qui utilise Python. Il offre une **large palette de fonctionnalités**, une **interopérabilité solide avec l'écosystème scientifique Python**, et une **bonne intégration avec des outils externes et des formats biologiques standards**. [3]

Exploration de Zenodo



I. Présentation de la plateforme Zenodo :

La plateforme **Zenodo** est un **réservoir de recherche en libre accès** créé **pour héberger, partager et préserver** les résultats de la recherche scientifique et les artefacts de recherche numérique. Elle a été **élaborée par l'Organisation européenne pour la recherche nucléaire (CERN)** dans le cadre du projet **OpenAIRE**, financé par l'Union européenne. [10]

Zenodo est ouverte à **toutes les disciplines scientifiques** et à tous les chercheurs, quel que soit leur pays ou leur institution. [10]

II. Objectifs principaux de Zenodo

Zenodo poursuit plusieurs objectifs clés :

1. Faciliter le partage et l'accès libre à la recherche

Zenodo permet aux chercheurs de **déposer leurs résultats, données, logiciels et autres artefacts de recherche**, de manière **ouverte et gratuite** pour toute la communauté scientifique. . [10]



2. Rendre les ressources de recherche trouvables et citables

Chaque contenu déposé sur Zenodo peut recevoir un **identifiant pérenne (DOI)**, ce qui facilite sa **citation dans les publications scientifiques** et sa traçabilité dans le temps. [11]



3. Assurer la conservation à long terme des objets de recherche

Les données et fichiers sont **stockés de façon fiable**, avec des sauvegardes et une infrastructure robuste fournie par le CERN. [11]

4. Faciliter le respect des exigences des politiques de science ouverte

Zenodo aide les chercheurs, notamment ceux soutenus par des programmes européens comme **Horizon Europe**, à se conformer aux **obligations d'accès ouvert et de réutilisation des données (FAIR)**. [10]

III. Types de contenus hébergés sur Zenodo

Zenodo accepte une **grande variété de contenus scientifiques**, notamment :

- **Publications scientifiques** : articles, rapports, prépublications
- **Données de recherche** : jeux de données brutes ou traitées
- **Logiciels et codes sources**
- **Images, vidéos et contenus multimédias scientifiques**
- **Documents de conférence, posters, présentations**
- **Documents méthodologiques, protocoles, outils numériques** [11]

IV. Intérêt de Zenodo pour la science ouverte et la recherche

1. Accès libre et diffusion mondiale

Zenodo permet à la communauté scientifique et **au grand public** d'accéder gratuitement à des résultats de recherche, favorisant ainsi la **diffusion des connaissances sans barrières financières**. [10]

2. Amélioration de la transparence scientifique

En déposant non seulement les articles publiés mais aussi les **données, codes et méthodes**, les chercheurs rendent leurs travaux plus **transparents et vérifiables** par d'autres scientifiques. [11]

3. Respect des principes FAIR

Zenodo facilite la mise en œuvre des principes **FAIR** (trouvé facilement, accessible, interopérable, réutilisable), ce qui est devenu une exigence dans beaucoup de projets de science ouverte et de financements européens. . [10]

4. Partage interdisciplinaire et collaboration

La plateforme n'est pas limitée à une seule discipline : elle permet le **croisement de données et de résultats entre domaines scientifiques différents**, ce qui est crucial pour les sciences de la nature et de la vie, où des données issues de différentes sources sont souvent combinées pour des analyses complexes. [10]

V. Pourquoi Zenodo est utile pour les sciences de la nature et de la vie

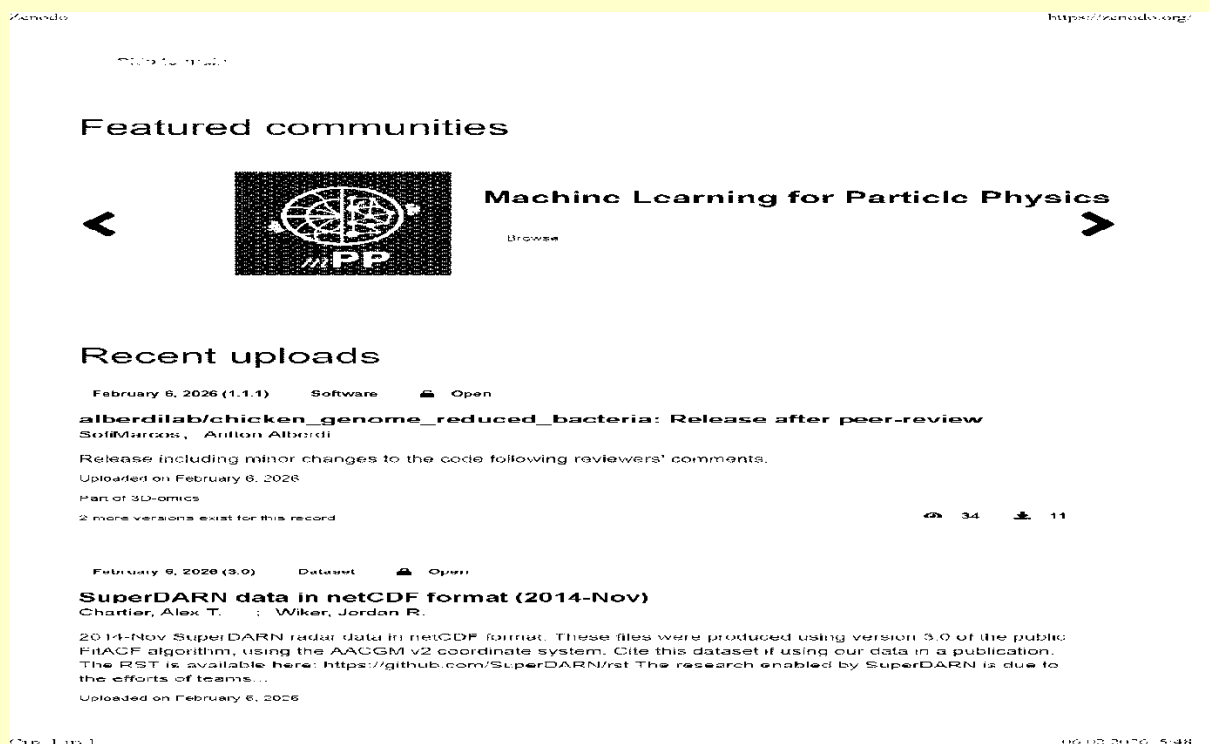
Pour les disciplines comme la **biologie, l'écologie, la géoscience ou la santé** :

- Zenodo permet **de partager des jeux de données expérimentaux**, facilitant ainsi leur réutilisation par d'autres chercheurs. [11]
- Les logiciels et scripts d'analyse (par exemple pour traiter des données génomiques ou écologiques) peuvent être **déposés et cités**, ce qui augmente la **fiabilité et la répliquabilité des résultats**. [11]
- Les résultats expérimentaux accompagnés de **métadonnées claires** sont plus faciles à retrouver et à combiner avec d'autres jeux de données. [11]

VI. Description des étapes réalisées :

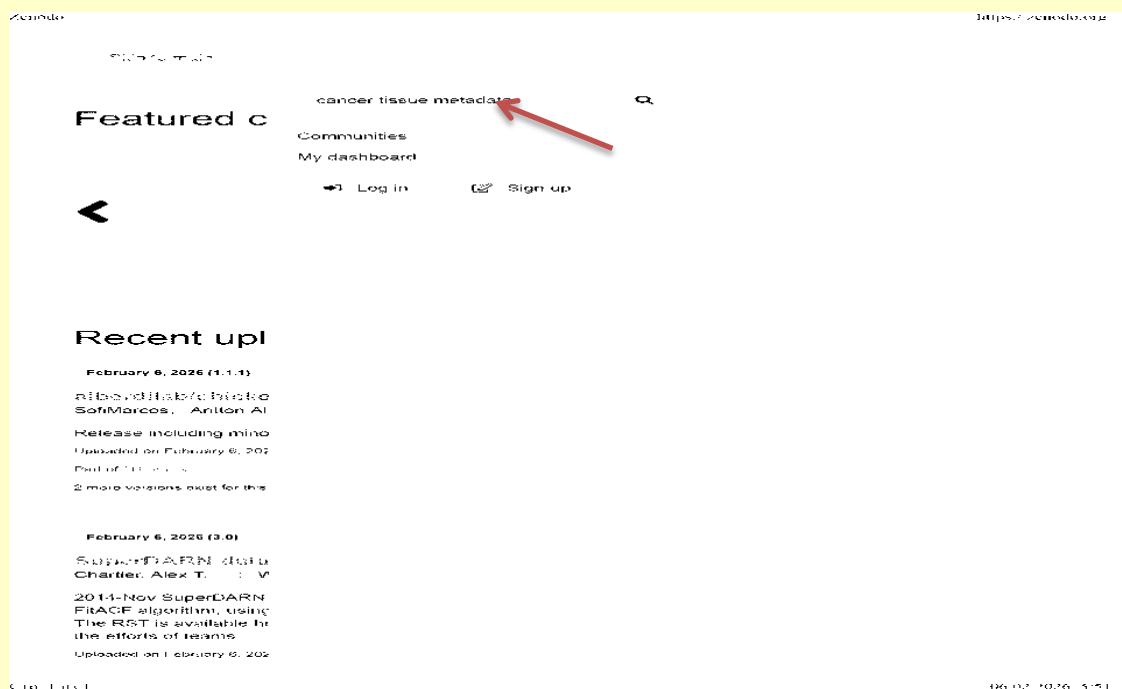
- **Recherche effectuée** : cancer tissue metadata
- **Critères de sélection de metadata** :
 - Taille de fichier : raisonnable (entre 1 Mo et 100 Mo)
 - Format de fichier : csv ,xml
 - Type de fichier : dataset
 - Statut d'accès : open
- **Navigation sur la plateforme** :

1. Accéder à Zenodo : Allez sur **zenodo.org**



2. Rechercher du contenu

- Utilisez la **barre de recherche** en haut de la page et écrivez (**cancer tissue metadata**)



- Filtres disponibles sur la gauche :
 - Type de document (**dataset**)
 - Taille de fichier : raisonnable (**entre 1 Mo et 100 Mo**)

- Format de fichier : **cvc ,xml**
- Type de fichier : **dataset**
- Statut d'accès : **open**
- Date de publication : **18 mai 2022**
- Accès libre

Search results

https://zenodo.org/search/?q=correct%20issue%20metadata&f=accession

10

Versions

View all versions

Access status

Open 123,320

Restricted 5,760

Embargoed 338

Resource types

Publication 70,900

Dataset 21,724

Image 15,818

Software 7,565

Presentation 4,107

Poster 1,919

Other 1,751

Video/Audio 368

Model 337

Lesson 327

Subjects

Sort by Best match

Network Fairly Represented in Single-Cell

Curated Human Cell Atlas single-cell RNA-seq sample reference raw - United Nations - based global population

Pre-emptive (unpublished) prior to 2021

Tissue microbiome composition and functional potential of histologically normal breast tissue (healthy, H) with those...

© 2025

ghts By Technology (In Situ Hybridization, Primary and Special Staining, Anatomic Pathology), By...

- Cliquez sur un résultat pour voir la **page détaillée** (description, auteurs, DOI, fichiers).


Composition and functional potential of the mammary microbiota prior to and following breast tumor diagnosis

https://zenodo.org/records/6561157

Stemsmma, Leah

Dryad

Published May 18, 2022 | Version v1

Dataset  Open

Composition and functional potential of the mammary microbiota prior to and following breast tumor diagnosis

Stemsmma, Leah

Show affiliations

These data were generated to analyze the human breast tissue microbiome composition and functional potential. We analyzed and compared the resident bacterial taxa of histologically normal breast tissue (healthy, H) with those of tissue donated prior to (pre-diagnostic, PD,) and after (adjacent normal, AN, and tumor, T) breast cancer diagnosis to determine the role of the breast tissue microbiome in breast cancer. DNA was isolated from 16S tissue samples, 9 negative controls, and 1 positive control and submitted for Illumina Miseq paired-end sequencing of the V3 - V4 region of the 16S gene. 16S amplicons were pre-processed into amplicon sequence variants (ASVs) using DADA2. To infer bacterial function in breast cancer, we predicted the functional bacteriome from these ASVs using PICRUSt2 and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.

Notes

Funding provided by U.S. Department of Defense Breast Cancer Research Priority Program (DPRP) Grant ID: W81XWH-20-2-0005. Award Number: W81XWH-20-2-0005.

Files

Attachment 1_6561157_1

Metadata used for Table 1

3. Consulter un dépôt

- Téléchargement des fichiers

- Citation prête à l'emploi
- Historique des versions
- Statistiques (vues, téléchargements)

VII. Métadonnées du dataset

- **Les métadonnées récupérer :**

```
<?xml version="1.0" encoding="utf-8" ?>
<= <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/
    http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
  <dc:creator>Stiemsma, Leah</dc:creator>
  <dc:date>2022-05-18</dc:date>
  <dc:description>&lt;p&gt;These data were generated to analyze the human breast
    tissue microbiome composition and functional potential. We analyzed and
    compared the resident bacterial taxa of histologically normal breast tissue
    (healthy, H) with those of tissues donated prior to (pre-diagnostic, PD, ) and
    after (adjacent normal, AN, and tumor, T) breast cancer diagnosis to determine
    the role of the breast tissue microbiome in breast cancer. DNA was isolated
    from 165 tissue samples, 9 negative controls, and 1 positive control and
    submitted for Illumina Miseq paired-end sequencing of the V3 - V4 region of the
    16S gene. 16S amplicons were pre-processed into amplicon sequence variants
    (ASVs) using DADA2. To infer bacterial function in breast cancer, we predicted
    the functional bacteriome from these ASVs using PICRUSt2 and the Kyoto
    Encyclopedia of Genes and Genomes (KEGG)
    database. &lt;/p&gt;</dc:description>
  <dc:description>&lt;p&gt;Funding provided by: U.S. Department of
    Defense&lt;br&gt;Crossref Funder Registry ID:
    http://dx.doi.org/10.13039/100000005&lt;br&gt;Award Number:
    W81XWH1810749&lt;/p&gt;</dc:description>
  <dc:identifier>https://doi.org/10.5061/dryad.9s4mw6mjd</dc:identifier>
  <dc:identifier>oai:zenodo.org:6561157</dc:identifier>
  <dc:publisher>Zenodo</dc:publisher>
  <dc:relation>https://doi.org/10.5281/zenodo.6554666</dc:relation>
  <dc:relation>https://zenodo.org/communities/dryad/</dc:relation>
  <dc:rights>info:eu-repo/semantics/openAccess</dc:rights>
  <dc:rights>Creative Commons Zero v1.0 Universal</dc:rights>
  <dc:rights>https://creativecommons.org/publicdomain/zero/1.0/legalcode</dc:rights>
  <dc:subject>mammary microbiome</dc:subject>
  <dc:subject>breast tissue</dc:subject>
  <dc:subject>functional metagenome</dc:subject>
  <dc:title>Composition and functional potential of the mammary microbiota prior to
    and following breast tumor diagnosis</dc:title>
  <dc:type>info:eu-repo/semantics/other</dc:type>
</oai_dc:dc>
```

- **Organisation claire des données :**

Tableau 01 : organisation des données récupérées .

Élément	Valeur
Titre	Composition and functional potential of the mammary microbiota prior to and following breast tumor diagnosis
Créateur / Auteur	Stiemsma, Leah
Date de publication	18 mai 2022
Description	Analyse de la composition et du potentiel fonctionnel du microbiome du tissu mammaire humain. Comparaison entre tissus mammaires normaux (sains), pré-diagnostiques, adjacents normaux et tumoraux. Séquençage Illumina MiSeq du gène 16S (région V3-V4), traitement avec DADA2, prédiction fonctionnelle via PICRUST2 et base KEGG.
Financement	U.S. Department of Defense
Identifiant du financeur	Crossref Funder Registry ID : 10.13039/1000000005
Numéro de subvention	W81XWH1810749
Identifiant principal (DOI)	https://doi.org/10.5061/dryad.9s4mw6mjd
Identifiant OAI	oai:zenodo.org:6561157
Éditeur / Plateforme	Zenodo
Relations	https://doi.org/10.5281/zenodo.6554666
Communauté Zenodo	https://zenodo.org/communities/dryad/records?q=&l=list&p=1&size=10&sort=newest
Droits d'accès	Accès libre (Open Access)
Licence	Creative Commons Zero (CC0 1.0)
URL de la licence	https://creativecommons.org/publicdomain/zero/1.0/legalcode
Mots-clés / Sujets	mammary microbiome ; breast tissue ; functional metagenome
Type de ressource	Autre (info:eu-repo/semantics/other)

Référence :

[1]: <https://biopython.org/?utm>

[2]: https://biopython.org/docs/1.84/Tutorial/chapter_introduction.html?utm

[3]: <https://bio.dev/tools/biopython?utm>

[4]: <https://es.wikipedia.org/wiki/Biopython?utm>

[6]: <https://en-test.vrachi.name/biopython/?utm>

[7]: <https://en.wikipedia.org/wiki/Biopython?utm>

[8]: <https://www.numberanalytics.com/blog/ultimate-guide-biopython-intro-bioinformatics?utm>

[9]: https://www.reddit.com/r/bioinformatics/comments/1egwgdx/seeking_alternatives_to_biopython_which_libraries/?utm

[10]: <https://about.zenodo.org/?utm>

[11]: <https://www.openaire.eu/zenodo-guide?utm>