



- **Faculté** : science de la nature et de la vie.
- **Département** : microbiologie appliquée et sciences alimentaires.
- **Module** : logiciels libres et open source.

TP :

Logiciels libres et open source

Réalisé par :

- **Laib Rania**
- **Bouadjel Hana.**
- **Moussaoui Nawal.**

*Année Universitaire
2025/2026*

sommaire

Partie 1 : étude théorique d'un outil (Biopython)

I.	présentation générale de Biopython	P02
II.	les fonctionnalités principales	P02
III.	Aspet technique	P04
IV.	piont forts.....	P04
V.	limites et points faibles	P05
VI.	conclusion.....	P05

Partie 2 : étude pratique .

I.	presentation de zenodo	P05
II.	descreption des étapes réalisées.....	P05
III.	Métadonnées du dataset.....	P08

partie 1 : étude théorique d'un outil (Biopython)


I. présentation générale de Biopython :

Le projet Biopython est une association internationale de développeurs proposant des outils Python libres pour la biologie moléculaire computationnelle. Python est un langage orienté objet, interprété et flexible, de plus en plus utilisé dans le calcul scientifique. Le site officiel de Biopython fournit des modules, scripts et ressources destinés aux chercheurs en bioinformatique. L'objectif principal de Biopython est de faciliter l'utilisation de Python en bioinformatique en proposant des modules réutilisables de haute qualité.

Biopython permet notamment de lire plusieurs formats de fichiers biologiques (FASTA, GenBank, BLAST...), d'accéder à des services en ligne comme NCBI ou ExpASY, et d'interfacer différents programmes bioinformatiques. Ainsi, Biopython constitue un outil essentiel et largement utilisé dans le domaine de la bioinformatique moderne. (1)

II. les fonctionnalités principales :

1- Manipulation des séquences biologiques Biopython permet d'importer, de lire et de manipuler des séquences (ADN, ARN, protéines) à l'aide des objets Seq et SeqRecord, ce qui facilite l'exploitation des données biologiques.



ATCGCG
TAGCGC
CGCGAT
AUCGCG
IA

Lesson 5

Applications of Seq Module Functions

Complement
Reverse Complement
Transcription
Back Transcription
Translation

Transcription

DNA

RNA

Protein

Translation

BioPython: Bio.SeqRecord

```
import Bio.SeqIO
import sys
import gzip

# Check the input
if len(sys.argv) < 2:
    print >>sys.stderr, "Please provide a sequence file"
    sys.exit(1)

# Get the sequence filename
seqfilename = sys.argv[1]

# Open the FASTA file and iterate through its sequences
seqfile = gzip.open(seqfilename)
for seq_record in Bio.SeqIO.parse(seqfile, "uniprot-xml"):
    # Print out the various elements of the SeqRecord
    print "\n-----NEW SEQUENCE-----\n"
    print "seq_record.annotations\n\t", seq_record.annotations
    print "seq_record.features\n\t", seq_record.features
    print "seq_record.dbxrefs\n\t", seq_record.dbxrefs
    print "seq_record.format('fasta')\n\t", seq_record.format('fasta')
    break
seqfile.close()
```

10/12/2013 BCH524 - 2013 - Edwards 13

2- Parsers pour les formats standards La principale force de Biopython sont ses "parsers", des modules capables de lire et manipuler les formats de données biologiques les plus répandus comme FASTA, GenBank ou FASTQ, facilitant l'intégration dans des workflows d'analyse.

3- Accès aux bases de données en ligne Biopython offre des interfaces pour accéder à des données biologiques telles que celles de ExpASY ou PubMed, ce qui rend plus simple la récupération de données depuis des banques en ligne.

4- Outils d'analyse des séquences Le module Bio.SeqUtils inclut plusieurs outils analytiques tels que : Calcul du taux de GC Recherche de motifs Analyse des six cadres de lecture ces outils aident à extraire des informations biologiques pertinentes.

5- Traduction de séquences Biopython peut traduire des séquences nucléotidiques en séquences protéiques en substituant les codons par leurs acides aminés correspondants, ce qui est crucial pour l'étude des gènes.

6- Requêtes BLAST en ligne Le module Bio.Blast permet d'envoyer des requêtes BLAST vers les serveurs du NCBI, pour comparer une séquence donnée avec celles des banques de données et identifier des régions similaires.

Alignment Results Information Panel

Search: Find Next

db	Name	Description	e-value	start point	align length	%identity	Include
ref	NM_024426	Homo sapiens Wilms tumor 1 (W...	0	1	3023	100	<input type="checkbox"/>
ref	NM_024424	Homo sapiens Wilms tumor 1 (W...	0	1	3023	99	<input type="checkbox"/>
emb	X51630	Human Wilms tumor WT1 mRNA f...	0	1	3009	99	<input type="checkbox"/>
ref	XM_008002762	PREDICTED: Chlorocebus sabae...	0	3056	3040	96	<input type="checkbox"/>
ref	XM_005578275	PREDICTED: Macaca fascicularis ...	0	1	3045	96	<input type="checkbox"/>
ref	XM_004050892	PREDICTED: Gorilla gorilla gorilla...	0	1	2755	99	<input type="checkbox"/>
ref	XM_008002769	PREDICTED: Chlorocebus sabae...	0	3056	3040	96	<input type="checkbox"/>

>gi|309951095|ref|NM_024426.4|
Homo sapiens Wilms tumor 1 (WT1), transcript variant D, mRNA
NM_024426

Score = 5583.54 bits (3023), Expect = 0
Identities = 3023/3023 (100%), Gaps = 0/3023 (0%), Strand Plus/Plus

Query	7	AGCTGGGGTAAGGAGITCAAGGCAGCGCCACACCGGGGGCTCTCCGCAACCCGACCGC	66
Subject	1	AGCTGGGGTAAGGAGITCAAGGCAGCGCCACACCGGGGGCTCTCCGCAACCCGACCGC	60
Query	67	CTGTCCGCTCCCCCACTTCCCGCCCTCCCTCCCACTACTCATTACCCACCCACCCACC	126
Subject	61	CTGTCCGCTCCCCCACTTCCCGCCCTCCCTCCCACTACTCATTACCCACCCACCCACC	120

Reset Select All Add Complete Sequences to Workspace Only Add Aligned Parts

Total hits for all sequences: 100.

7- Extraction et traitement des données depuis Entrez Biopython communique avec les outils de recherche comme Entrez (NCBI), permettant d'extraire et de traiter automatiquement des informations depuis différentes bases de données (Gene, PubMed, etc).

8 - Alignement de séquences Biopython peut utiliser des programmes externes d'alignement tels que MUSCLE pour effectuer des alignements multiples de séquences à partir de fichiers FASTA. (2)

Cut & Paste input - FASTA

File Edit Select View Format Colour Calculate Web Service

	240	250	260	270
Groupe_A/1-1062	GTGGAAGGATGTCCTCGTGGTGACCCCTTGGCTGGCTCCCATTTGT			
Groupe_B/1-1062	GTGGAAGGATGTCCTCGTGGTGACCCCTTGGCTGGCTCCCATTTGT			
Groupe_O/1-1061	GTGGAAGGATGTCCTCGTGGT-AACCCCTTGGCTGGCTCCCATTTGT			

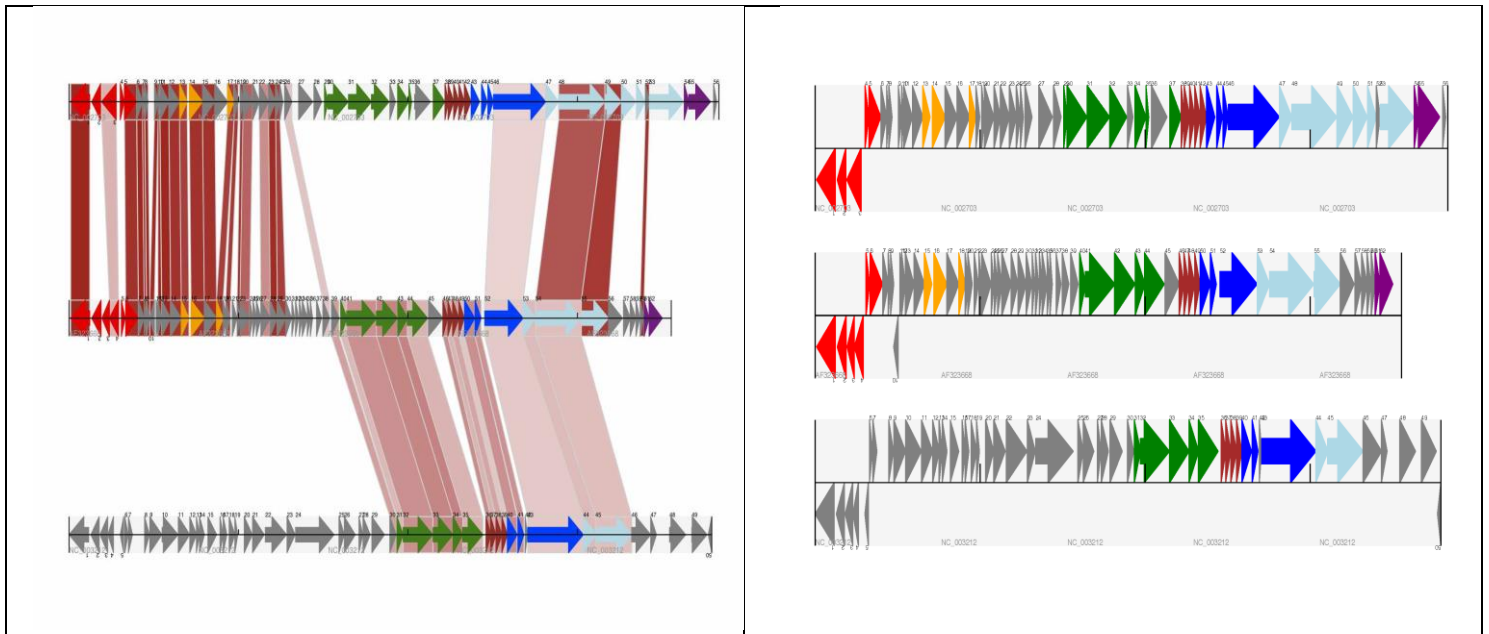
Consensus

GTGGAAGGATGTCCTCGTGGTGACCCCTTGGCTGGCTCCCATTTGT

Sequence position 237 G 100%

9- Analyse des structures macromoléculaires Parsing des fichiers PDB et manipulation des structures via Bio.PDB.

10- Représentation graphique Visualisation des génomes et des séquences avec GenomeDiagram.



11- Génétique des populations Simulation coalescente (SIMCOAL2), wrappers pour GENEPOP, détection de sélection (Fst-outlier) Bases de données relationnelles Stockage et récupération de séquences annotées dans SQL via BioSQL avec ORM pour SeqRecord. (3)

III. Aspet technique :

A-Téléchargement et Installation Pour Windows, des installateurs “click-and-run” sont fournis. La plupart des distributions Linux incluent un paquet Biopython optionnel (bien que celui-ci puisse être obsolète). Sinon, il est généralement nécessaire de télécharger et de décompresser l’archive, puis d’installer depuis la source. Pour plus de détails, y compris les prérequis, consultez la page de téléchargement officielle.

Exemple rapide et vérification de l’installation Après l’installation de Biopython, vous pouvez tester que tout fonctionne correctement avec un petit exemple Python. Par exemple : `from Bio.Seq import Seq` .

-**création d’un objet séquence** `my_seq = Seq("CATGTAGACTAG")` .

- **affichage de quelques détails** `print("seq %s is %i bases long" % (my_seq, len(my_seq))) print("reverse complement is %s" % my_seq.reverse_complement()) print("protein translation is %s" % my_seq.translate())`

- Sortie attendue : seq CATGTAGACTAG is 12 bases long reverse complement is CTAGTCTACATG protein translation is HVD. Cela montre rapidement le fonctionnement de l’objet Seq de Biopython et quelques méthodes principales associées.

B- Lecture et écriture de fichiers de séquences Biopython fournit le module SeqIO pour lire ou écrire des séquences sous forme d’objets SeqRecord. Pour les fichiers d’alignement multiple de séquences, le module AlignIO peut être utilisé comme alternative. Pour les débutants souhaitant apprendre Python et mieux comprendre le fonctionnement de Biopython : A Byte of Python Dive Into Python Python Quick Reference Le Biopython Tutorial (HTML, PDF) Article : Bassi 2007, PMID: 18052533 Vous pouvez également consulter le diagramme des classes pour comprendre les relations entre les modules, et utiliser les outils de recherche sur le Wiki pour obtenir plus d’informations sur des sujets spécifiques.

(4)

IV. piont forts :

Biopython est un projet open source et disponible gratuitement, ce qui facilite son utilisation et son accès. C'est un projet mature, fruit d'une collaboration internationale de développeurs bénévoles. Biopython fournit des bibliothèques Python pour un large éventail de problèmes bioinformatiques, y compris la lecture et l'écriture de différents formats de fichiers de séquences et d'alignements multiples. Il permet de travailler avec des structures macromoléculaires 3D. Il offre des interfaces avec des outils courants comme BLAST, ClustalW et EMBOSS, ainsi qu'un accès à des bases de données en ligne clés. Biopython inclut également des méthodes numériques pour l'apprentissage statistique. (5)

Biopython fournit une large gamme d'outils et de modules pour diverses tâches en bioinformatique. Facilité d'utilisation : L'API de Biopython est conçue pour être conviviale, ce qui la rend accessible aux chercheurs ayant différents niveaux d'expérience en programmation. Soutien de la communauté : Biopython dispose d'une communauté active, garantissant que la bibliothèque est régulièrement mise à jour et maintenue. (6)

V. limites et points faibles :

Biopython peut ne pas être la bibliothèque la plus efficace pour des calculs à très grande échelle. Support limité pour certains formats de fichiers : Bien que Biopython prenne en charge de nombreux formats de fichiers courants, il peut ne pas supporter des formats plus spécialisés ou propriétaires. (7)

VI. conclusion :

Biopython est une interface de programmation d'applications (API) open source étendue, utilisée à la fois dans le développement de logiciels en bioinformatique et dans les scripts quotidiens pour des tâches bioinformatiques courantes. Les fonctionnalités décrites ici ne représentent qu'un sous-ensemble ; les utilisateurs potentiels devraient se référer au tutoriel et à la documentation de l'API pour obtenir des informations complémentaires. (8)

Partie 2 : etude pratique

I. presentation de zenodo :

1. **Objectifs de la plateforme** : Zenodo est une plateforme en ligne gratuite conçue pour permettre aux chercheurs de stocker, partager et diffuser leurs résultats scientifiques. Son objectif principal est de faciliter l'accès ouvert aux données et publications et d'encourager la transparence et la collaboration dans la recherche.
2. **Types de contenus hébergés** : Zenodo héberge différents types de contenus scientifiques, notamment : Jeux de données (Datasets) provenant d'expériences ou de projets de recherche Articles scientifiques et rapports Logiciels et codes source pour la recherche Supports de présentation et posters Autres ressources scientifiques utiles pour la communauté
3. **Intérêt pour la science ouverte et la recherche en sciences de la nature et de la vie** : La plateforme favorise la science ouverte en offrant un accès libre et durable aux ressources scientifiques. Dans le domaine des sciences de la nature et de la vie, Zenodo permet : Une diffusion rapide et fiable des données expérimentales La réutilisation des données par d'autres chercheurs pour valider ou compléter des études L'attribution d'identifiants persistants (DOI) garantissant la traçabilité et la reconnaissance scientifique

II. description des étapes réalisées :

1. **Recherche effectuée (requête utilisée)** : Pour cette étude pratique, la recherche a été effectuée sur la plateforme Zenodo en utilisant le mot-clé “genome” pour trouver des datasets pertinents sur les génomes et les données biologiques associées.
2. **Critères de sélection du dataset** : Le dataset a été sélectionné selon les critères suivants :
 - 1_Type de ressource : Dataset (pas publication)
 - 2_Accessibilité : Open / libre téléchargement
- 3_Pertinence scientifique : Étude sur le génome de papillons Heliconius, correspondant au mot-clé utilisé
- Complétude des métadonnées : DOI, License, Keywords et Description disponibles.
3. **Navigation sur la plateforme** : La navigation sur Zenodo a été réalisée comme suit :
 - 1_Accès à <https://zenodo.org>.

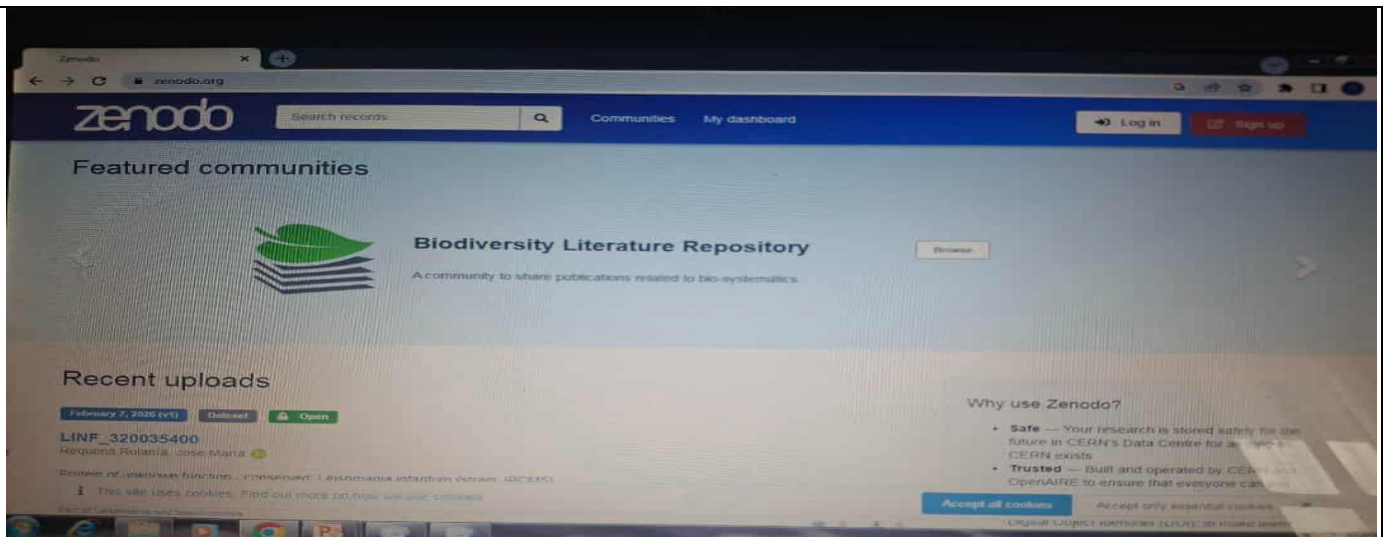


Figure 1 :page d'accueil de zenodo.

2_Entrée du mot-clé “genome” dans la barre de recherche.

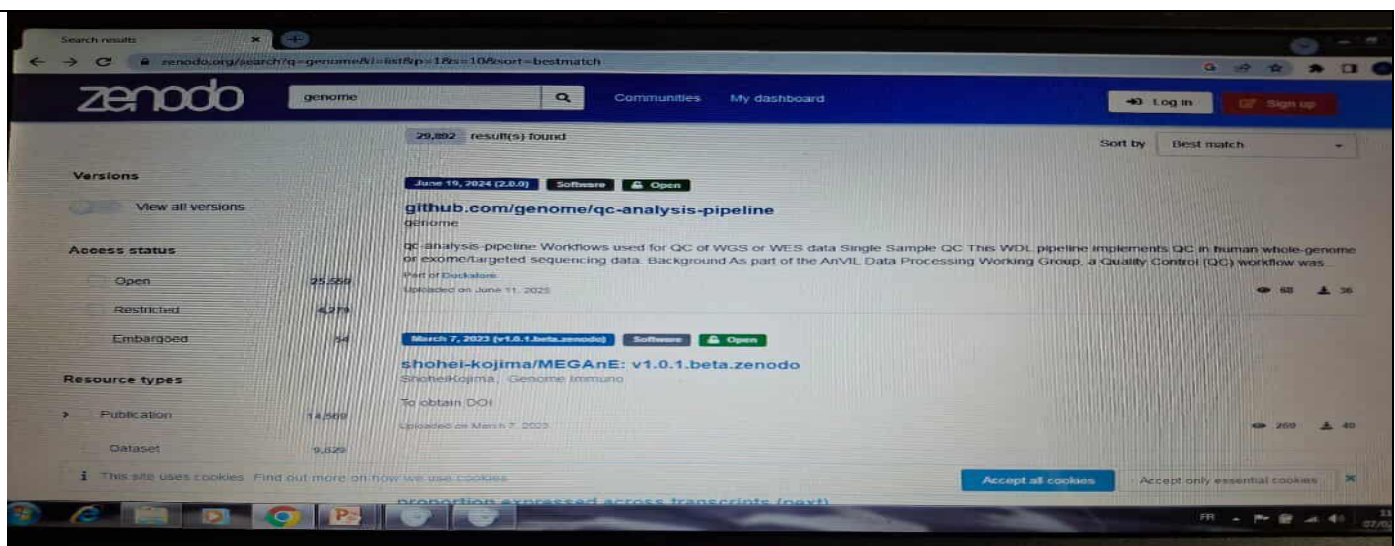


Figure 2 : résultats de recherche avec le mot-clé « génome ».

3_Filtrage des résultats par Dataset (ressource).

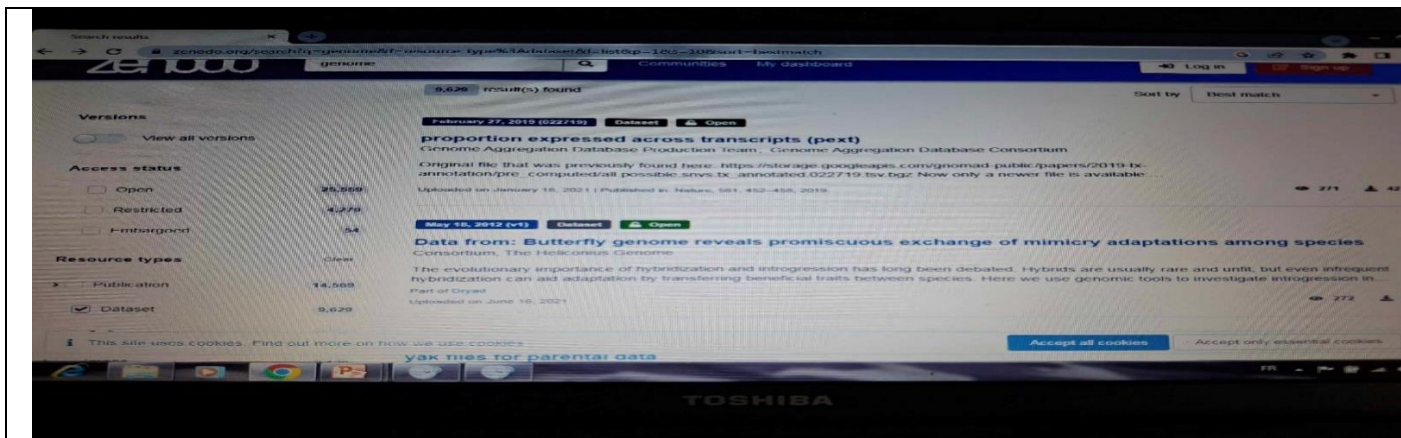


Figure 3 : filtrage des résultats pour afficher uniquement les datasets.

4_Sélection du dataset pertinent et consultation de sa page.

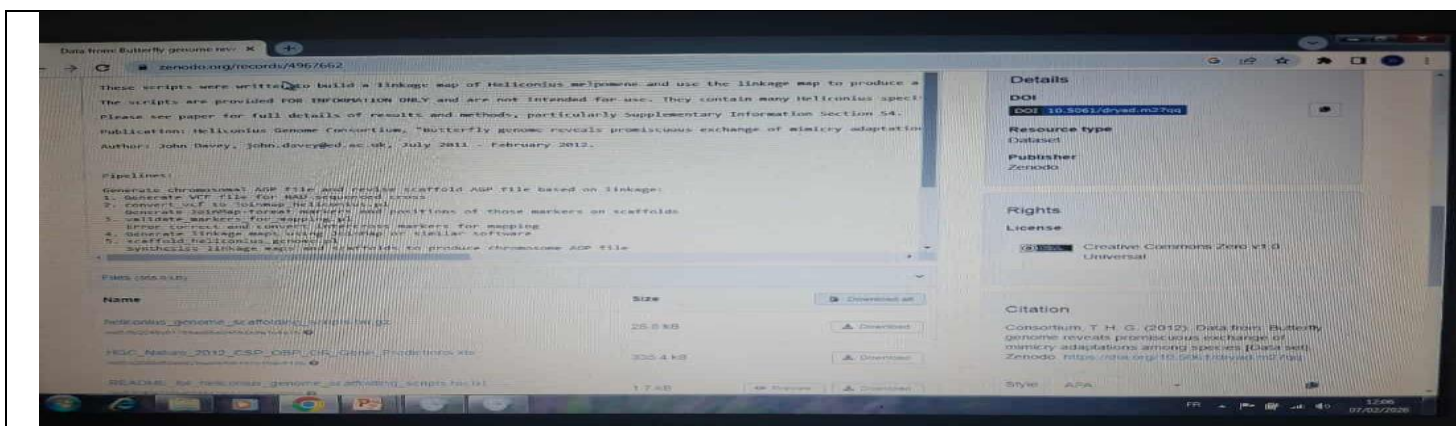


Figure 4 : page du dataset sélectionné sur zenodo.

5_Téléchargement du dataset via le bouton "Download all" et récupération des métadonnées dans les sections Details et Rights.

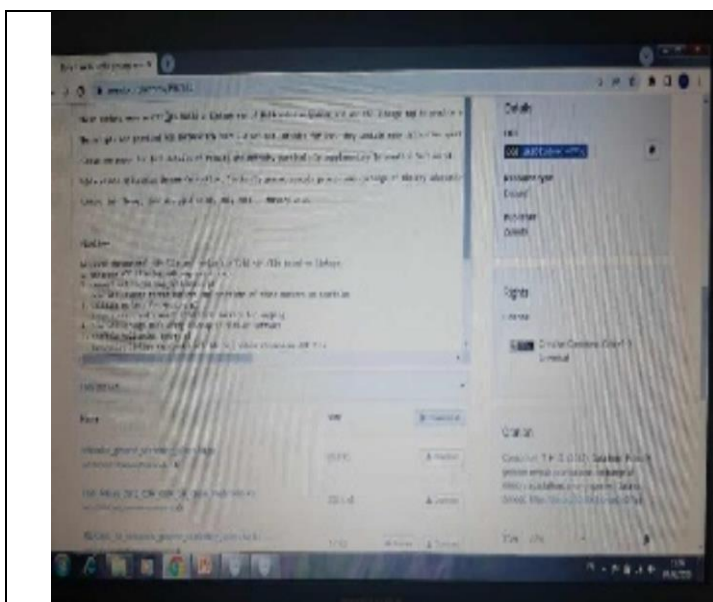


Figure 5 : téléchargement du dataset depuis zenodo.

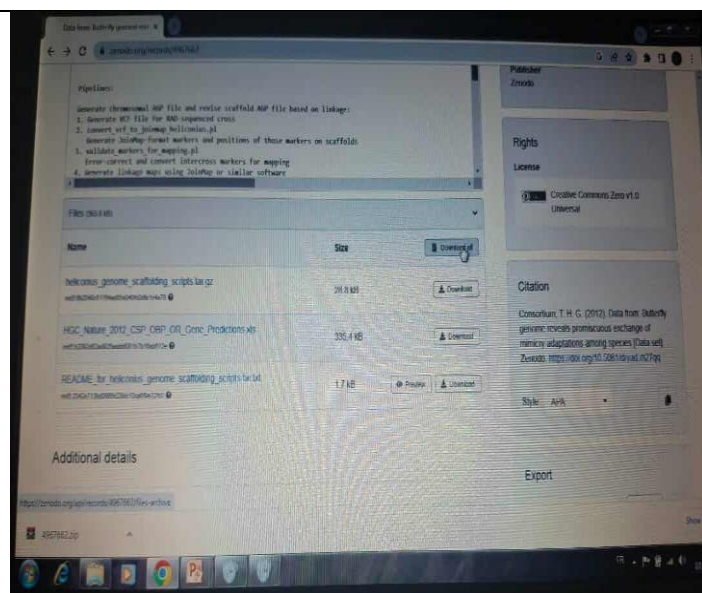


Figure 6 : métadonnées principales du dataset.

III. Métadonnées du dataset:

A_Présentation structurée des métadonnées récupérées : Les informations du dataset ont été récupérées directement depuis les sections “Details” et “Rights” de la page Zenodo. Ces métadonnées permettent de décrire le dataset de manière claire et complète selon la norme Dublin Core.

B_Organisation claire (tableau recommandé) : Le tableau ci-dessous présente de manière structurée toutes les métadonnées récupérées pour le dataset sélectionné, conformément à la norme Dublin Core.

Elément dublin core	Information extraite du dataset
Titre (Title)	Data from: Butterfly genome reveals promiscuous exchange of mimicry adaptations among species
Créateur (Creator)	The heliconius genome consortium
Date de publication (Date)	18 mai 2012
Type de ressource	dataset
Éditeur	zenodo
Identifiant (Identifier / DOI)	10.5061/dryad.m27qq
Lien DOI complet	https://doi.org/10.5061/dryad.m27qq
Licence (Rights / License)	Creative commons zero V1.0 Universal (CCO)
Communauté	dryad
Mots-clés (Keywords)	Scaffolding ;linkage mapping ; heliconius melpomene
Description	Ce dataset contient des données génomiques sur les papillons du genre Heliconius, utilisées pour étudier l'échange de gènes liés aux adaptations de mimétisme entre différentes espèces.
Citation recommandée	Consortium, T.H.G. (2012). Data from: Butterfly genome reveals promiscuous exchange of mimicry adaptations among species [Data set]. Zenodo. https://doi.org/10.5061/dryad.m27qq .

➤ **Refirance :**

- (1) : .org/Biopython Documentation, <https://biopython.org>.
- (2) : Biopython: freely available Python tools for computational molecular biology and bioinformatics Peter J A Cock 1,*, Tiago Antao 2, Jeffrey T Chang 3, Brad A Chapman 4, Cymon J Cox 5, Andrew Dalke 6, Iddo Friedberg 7, Thomas Hamelryck 8, Frank Kauff 9, Bartek Wilczynski 10,11, Michiel J L de Hoon 12.
- (3) : Tutoriel et livre de recettes Biopython Jeff Chang, Brad Chapman, Iddo Friedberg, Thomas Hamelryck, Michiel de Hoon, Peter Cock, Tiago Antao, Eric Talevich, Bartek Wilczyński Dernière mise à jour – 10 janvier 2024 (Biopython 1.8 .
- (4) : The Biopython Project. Download and Installation. <https://biopython.org/wiki/Download>.
- (5) : Biopython: freely available Python tools for computational molecular biology and bioinformatics Peter J A Cock et al. Bioinformatics. 2009.
- (6) : Lee, Sarah. Unlocking Biopython's Potential: A Comprehensive Guide to Bioinformatics with Python. AI generated, Llama-4-Maverick-17B-128E-Instruct-FP8, 19 juin 2025.
- (7) : 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. Bioinformatics, 25(11), 1422–1423.
- (8) : Biopython: freely available Python tools for computational molecular biology and bioinformatics Peter J A Cock et al. Bioinformatics. 2009.