

**République Algérien Démocratique et populaire**  
**Université Mohamed Sadik ben Yahia-Jijel Faculté**  
**des sciences de la nature et de la vie**

**Département science de l'environnement et des  
sciences agronomique**



***Rapport Devoir de Logiciel***

**Propose par :**

**\*DR.Bensalem**

**Réalise par :**

**\*Zitouni Nada**

**\*Frites Samar**

**\*Bazniar Amel**

**Année : 2025/2026**

## ***Introduction***

Dans le contexte actuel de la recherche scientifique, logiciels libres et open source occupent une place dans le développement des sciences de la nature et de la vie. Ils permettent aux chercheurs d'accéder à des puissants, flexibles et reproductibles, tout en favorisant collaboration et le partage des connaissances à internationale

L'utilisation de ces outils contribue à améliorer transparence scientifique, la reproductibilité des résultats l'innovation technologique. En bioinformatique, génomique ou encore en modélisation biologique, nombreuses solutions open source sont indispensables pour analyser et interpréter les grandes quantités de données biologiques générées

Ce travail s'inscrit dans cette dynamique et vise à d'une part, un outil open source largement utilisé en biologiques, et d'autre part, à explorer la plateforme Zenodo, qui joue un rôle central dans le partage des données scientifiques en accès libre. À travers cette étude et pratique, nous chercherons à comprendre des logiciels libres dans la recherche moderne ainsi que leur contribution à la science

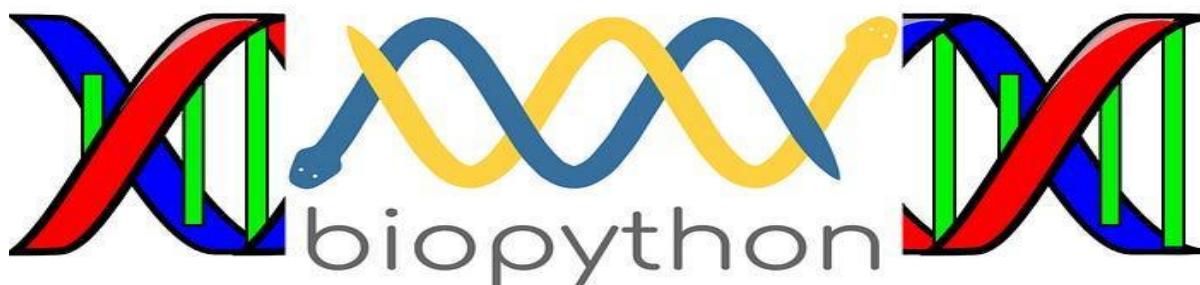
## PARTIE 1: Étude théorique

Outil choisi :

### 1-Présentation générale de

**Biopython** est une bibliothèque open source écrite en langage Python, destinée au traitement et à l'analyse données biologiques.

Elle a été créée en 1999 et est développée par une communauté internationale de chercheurs et développeurs.



L'objectif principal de **Biopython** est de fournir des simples et puissants pour la bioinformatique, pour :

- La manipulation de séquences biologiques (ADN, protéines).
- L'accès aux bases de données biologiques.
- L'analyse

-Le traitement de fichiers biologiques

**Biopython** est largement utilisée dans la recherche génomique, transcriptomique et protéomique.

## 2-Fonctionnalités

**Biopython** offre plusieurs modules

### a) Manipulation des séquences

-Lecture et écriture de fichiers FASTA, GenBank, EMBL

Traduction ADN →

-Calcul de composition en

-Alignement de séquences

### b) Accès aux bases de données biologiques :

-Connexion à NCBI (GenBank, PubMed...)

-Téléchargement automatique de

### c) Analyse

-Construction et visualisation d'arbres

### d) Bioinformatique

-Lecture de fichiers PDB (structures

### e) Outils complémentaires :

-Analyse de

-Statistiques

-Interfaces avec d'autres logiciels (ClustalW, BLAST...)

### 3-Aspects

**Langage :** Python

**Licence :** Open source (Biopython License)

**Systèmes compatibles :** Windows, Linux,

**Dépendances :** nécessite Python (version 3.x recommandée)

**Installation :** via

### Exemple

pip install biopython

Biopython est modulaire et s'intègre facilement dans des scripts Python ou des projets de data science.

### 4-Points forts :

\*Gratuit et open source.

\*Large communauté et documentation officielle

\*Compatible avec de nombreux formats

\*Facile à utiliser pour les étudiants et

\*S'intègre bien avec d'autres bibliothèques Python (NumPy, pandas, matplotlib).

### 5-Limites et points faibles :

- \* Nécessite des connaissances en programmation
- \* Peut être lent pour les très grands datasets.
- \* Certaines fonctionnalités avancées nécessitent des externes.
- \* Documentation parfois technique pour les

## **6-Conclusion :**

Biopython est un outil puissant et flexible pour bioinformatique

Grâce à sa nature open source et sa grande communauté, il constitue une solution idéale pour l'analyse des biologiques dans le domaine des sciences de la nature et de la vie.

Il est particulièrement adapté aux projets de académique et à l'enseignement de la

## **Références :**

<https://biopython.org>

<https://en.wikipedia.org/wiki/Biopython>

## **PARTIE 2 :Exploration de**

### **1. Présentation de**

#### **1.1 Objectifs de la**

Zenodo est une plateforme de dépôt et de diffusion en accès, créée en 2013 par le CERN avec le soutien de Commission européenne. Elle s'inscrit dans la dynamique de la science ouverte, en offrant aux chercheurs un espace pour archiver, partager et valoriser leurs scientifiques

L'un des objectifs majeurs de Zenodo est d'assurer l'accessibilité, la pérennité et la citabilité des résultats de recherche, grâce à l'attribution systématique d'un numérique permanent (DOI) à chaque dépôt.

#### **1.2 Types de contenus hébergés :**

La plateforme Zenodo permet l'hébergement de divers types de productions scientifiques, couvrant l'ensemble disciplines de recherche. Parmi les contenus les plus fréquemment déposés, on trouve :

- des jeux de données scientifiques
- des articles et
- des logiciels et codes sources.
- des figures, images et visualisations

-des présentations, posters et rapports

Cette diversité de contenus rend [Zenodo](#) adaptée aux besoins des chercheurs et des étudiants de niveau master et doctorat.

### **1.3 Intérêt de Zenodo pour la science ouverte et les sciences de la nature et de la vie :**

Dans le cadre de la science ouverte, [Zenodo](#) constitue un moyen stratégique pour la diffusion et la réutilisation des données de recherche. En sciences de la nature et de la vie, la plateforme facilite le partage de données expérimentales issues de travaux en biologie, écologie, génomique ou sciences de l'environnement.

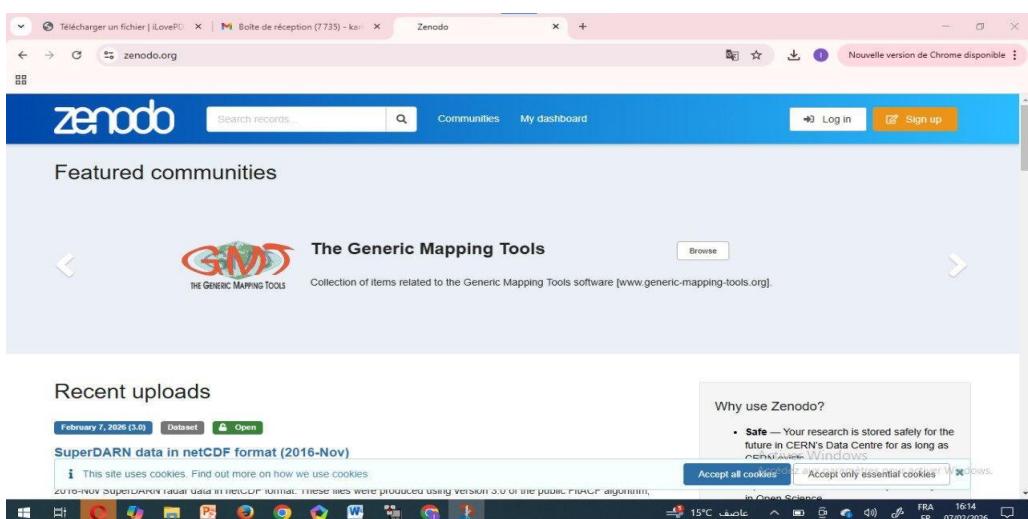
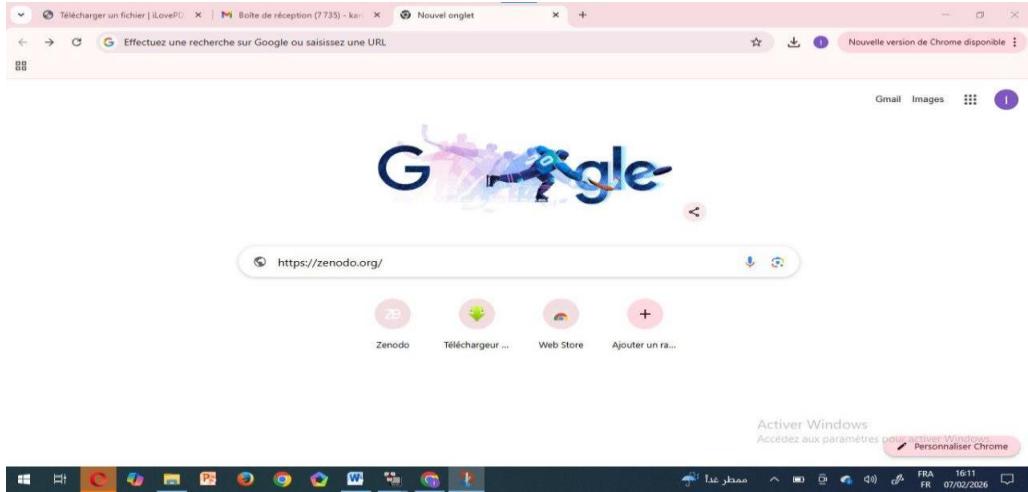
L'accès libre à ces données contribue à renforcer la transparence scientifique, la reproductibilité des résultats et la collaboration interdisciplinaire. De plus, [Zenodo](#) garantit la conservation à long terme des données, répondant ainsi aux exigences actuelles de gestion et de valorisation des données de recherche (Research Data Management).

## **2. Description des étapes**

### **2.1 Recherche effectuée sur la plateforme**

La première étape a consisté à accéder à la plateforme [Zenodo](#) via le site officiel

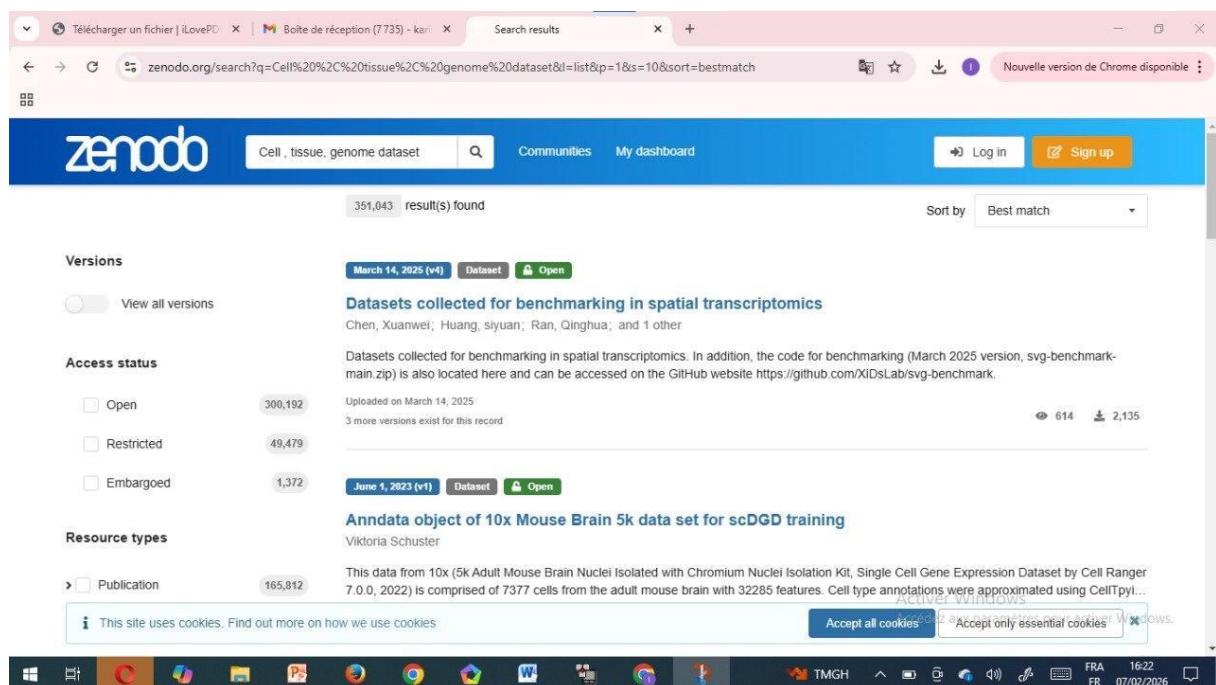
). Une recherche a ensuite été effectuée à l'aide du moteur de recherche intégré à la plateforme.



## Commentaire :

Figure : Page d'accueil de Zenodo, montrant le moteur de recherche et les options principales de

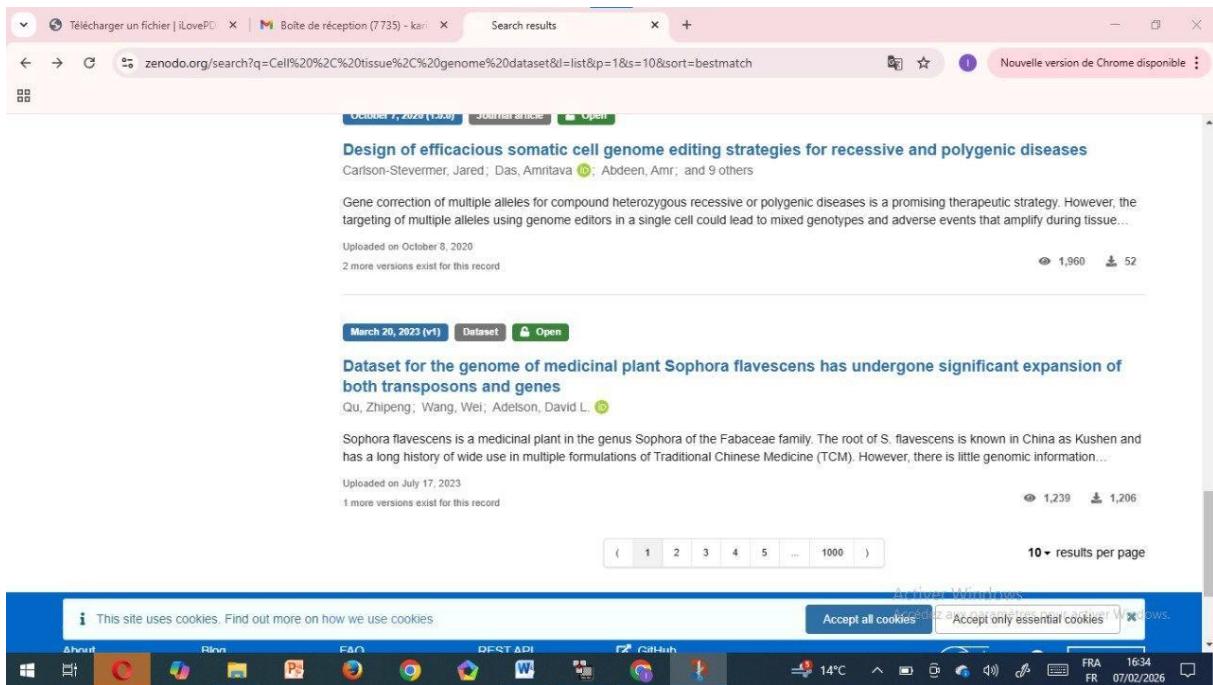
La requête utilisée est le mot-clé «cell, tissu, genome», en raison de sa pertinence pour les sciences de la vie et lien direct avec les données biologiques et génomiques.



## Commentaire :

Figure : Résultats obtenus après la recherche avec le mot- « cell , tissu , genome

Les résultats obtenus présentent plusieurs types de ressources, notamment des jeux de données, des publications scientifiques et des fichiers associés à des de recherche.



## Commentaire :

Figure : Résultats obtenus Chaque entrée affiche le l'auteur, la date et le type de contenu.

## 2.2 Critères de sélection du

Parmi les résultats proposés, un jeu de données a été sélectionné sur la base de plusieurs critères scientifiques techniques, à savoir :

-la pertinence du dataset par rapport au domaine sciences de la nature et de la vie.

-la clarté de la description fournie par la disponibilité des métadonnées associées.

-la présence d'un identifiant DOI garantissant la traçabilité jeu de données.

-le format des fichiers, permettant une scientifique

Ces critères ont permis de choisir un dataset fiable, bien documenté et conforme aux principes de la science ouverte

The screenshot shows a web browser window with multiple tabs open. The active tab is titled "Dataset for the genome of medicinal plant Sophora flavescens has undergone significant expansion of both transposons and genes". The URL in the address bar is zenodo.org/records/8153260. The page header includes the Zenodo logo, a search bar, and links for "Communities" and "My dashboard". A "Dataset" button and an "Open" button are visible. Below the header, it says "Published March 20, 2023 | Version v1". The main content area features a summary of the dataset: "Dataset for the genome of medicinal plant Sophora flavescens has undergone significant expansion of both transposons and genes" by Qu, Zhipeng<sup>1</sup>; Wang, Wei<sup>2</sup>; Adelson, David L.<sup>1</sup>. It includes a "Show affiliations" link. A detailed description follows, mentioning the plant's history and the study's methods. To the right, there are statistics: "1K VIEWS" and "1K DOWNLOADS", with a "Show more details" link. A "Versions" section lists two versions: "Version v1" (Mar 20, 2023) and "Version v1" (Mar 20, 2023). A "View all 2 versions" link is also present. At the bottom, a cookie consent banner asks if the user wants to "Accept all cookies" or "Accept only essential cookies". The browser interface at the bottom shows various icons and system status information.

## Commentaire :

Figure : Page détaillée du dataset choisi. On peut consulter la description, télécharger les fichiers et accéder métadonnées.

### 2.3 Navigation et utilisation de la

La navigation sur la plateforme **Zenodo** est intuitive structurée. Chaque jeu de données dispose d'une page dédiée regroupant :

-une description détaillée du

-les fichiers téléchargeables.

-les informations relatives aux

-les métadonnées normalisées.

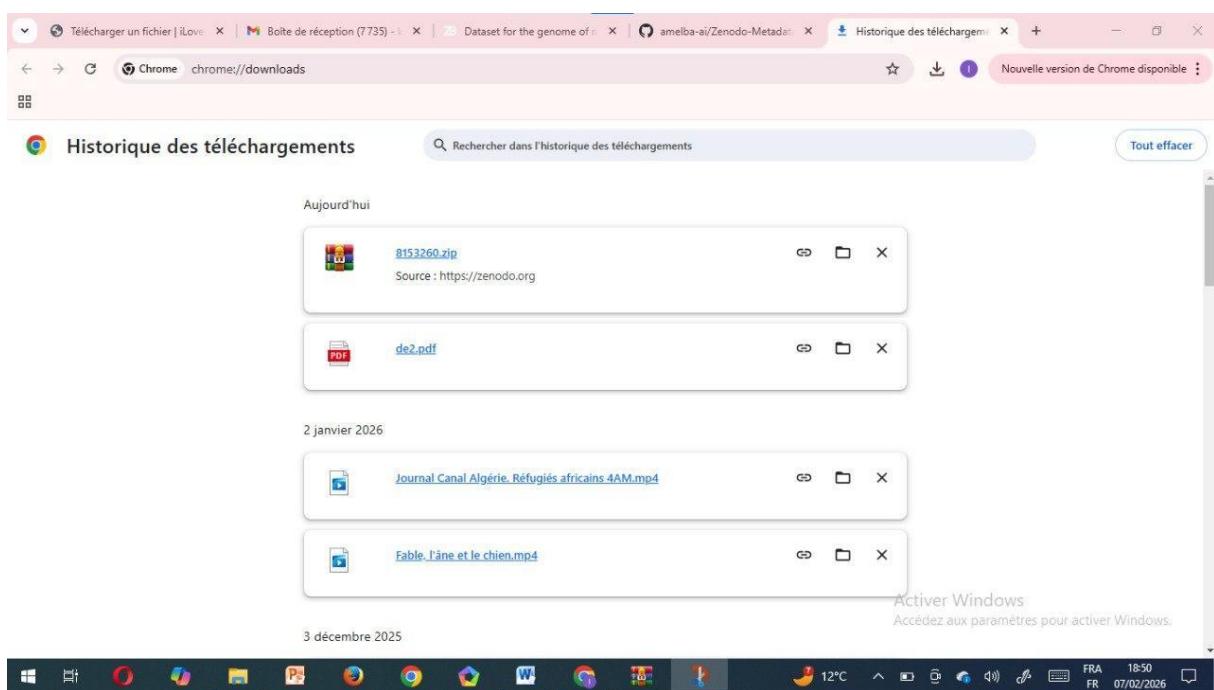
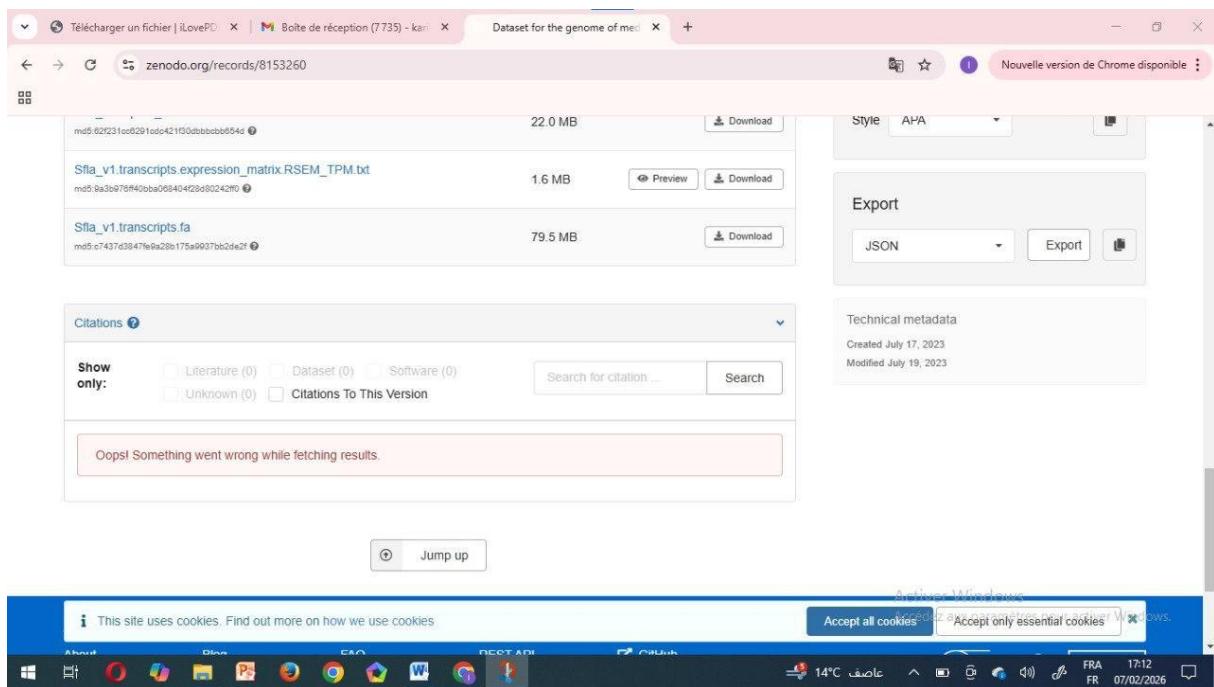
-la licence d'utilisation et le

Après la sélection du dataset, les fichiers associés ont téléchargés à l'aide de l'option Download. Les ont ensuite été consultées directement sur la page du dépôt afin d'être exploitées selon une norme de standardisée

The screenshot shows a web browser window with the following details:

- Address Bar:** zenodo.org/records/8153260
- Title Bar:** Télécharger un fichier | iLovePD | Boîte de réception (7735) - kain | Dataset for the genome of medicin... | Nouvelle version de Chrome disponible
- Content Area:**
  - Files (4.9 GB):** A table listing seven files:

Name	Size	Action
Sfla_v1.cdna.gff3	71.2 MB	<a href="#">Download</a>
Sfla_v1.chromosomes.fa	2.1 GB	<a href="#">Download</a>
Sfla_v1.proteins.fa	26.1 MB	<a href="#">Download</a>
Sfla_v1.repeat_EDTA.chromosomes_softMasked.fa	2.1 GB	<a href="#">Download</a>
Sfla_v1.repeat_EDTA.gff3	522.4 MB	<a href="#">Download</a>
Sfla_v1.repeat_EDTA.lib	22.0 MB	<a href="#">Download</a>
Sfla_v1.transcripts.expression_matrix.RSEM TPM.txt	1.6 MB	<a href="#">Preview</a> <a href="#">Download</a>
  - Rights:** Creative Commons Attribution 4.0 International
  - Citation:** Qu, Z., Wang, W., & Adelson, D. L. (2023). Dataset for the genome of medicinal plant Sophora flavescens has undergone significant expansion of both transposons and genes (Version v1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.8153260
  - Style:** APA
  - Cookie Consent:** This site uses cookies. Find out more on how we use cookies. Accept all cookies Accept only essential cookies
- Bottom Bar:** Windows taskbar showing various pinned icons and system status (14°C, 17:12, 07/02/2026)



## Commentaire :

Figure : Téléchargement du fichier dataset depuis prêt à être utilisé pour

### 3. Métadonnées du dataset

Tableau des métadonnées (Dublin Core)

Élément Valeur

Title	Dataset for the genome of medicinal plant Sophora flavescens has undergone expansion of both transposons and genes
Creator	Qu, Zhipeng ; Wang, Wei2 ;Adelson, David L.1
Subject	WGS; nanopore; Sophora ;Genome assembly; Medicinal plant
Description	Sophora flavescens is a medicinal plant in the genus Sophora of the Fabaceae family. The root of <i>S. flavescens</i> is known in China as Kushen and has a long history of wide use in formulations of Traditional Chinese (TCM). However, there is little information available for <i>S. flavescens</i> , has greatly hindered the breeding of <i>S. flavescens</i> and characterisation of compounds. Therefore, in this study, we used third-generation Nanopore long-sequencing technology combined with Hi-C scaffolding technology to de novo assemble <i>S. flavescens</i> genome. We obtained chromosomal level high-quality <i>S. flavescens</i> draft genome. The draft genome size approximately 2.08 Gb, with more than 80% annotated as Transposable Elements (TEs). We also annotated 60,485 genes and examined their expression profiles in leaf, stem and root

	tissues. We also characterised the genes pathways involved in the biosynthesis of major bioactive compounds, including flavonoids and isoflavonoids. The genome provides valuable resources for conservation, genetic research and breeding <i>S. flavescens</i> .
Publisher	Zenodo
Date	March 20, 2023
Format	CSV
Identifie	10.5281/zenodo.8153260
Language	English
Rights / License	CC BY 4.0

The screenshot shows a Zenodo dataset page for the genome of medicinal plant *Sophora flavescens*. The page includes the following information:

- Title:** Dataset for the genome of medicinal plant *Sophora flavescens* has undergone significant expansion of both transposons and genes
- Authors:** Qu, Zhipeng<sup>1</sup>; Wang, Wei<sup>2</sup>; Adelson, David L.<sup>1</sup>
- Abstract:** *Sophora flavescens* is a medicinal plant in the genus *Sophora* of the Fabaceae family. The root of *S. flavescens* is known in China as Kushen and has a long history of wide use in multiple formulations of Traditional Chinese Medicine (TCM). However, there is little genomic information available for *S. flavescens*, which has greatly hindered the breeding of *S. flavescens* and characterisation of bioactive compounds. Therefore, in this study, we used third-generation Nanopore long-read sequencing technology combined with Hi-C scaffolding technology to *de novo* assemble the *S. flavescens* genome. We obtained a chromosomal level high-quality *S. flavescens* draft genome. The draft genome size is approximately 2.08 Gb, with more than 80% annotated as Transposable Elements (TEs). We also annotated 60,485 genes and examined their expression profiles in leaf, stem and root tissues. We also characterised the genes and pathways involved in the biosynthesis of major bioactive compounds, including alkaloids, flavonoids and isoflavonoids. The assembled genome provides valuable resources for conservation, genetic research and breeding of *S. flavescens*.
- Statistics:** 1K VIEWS, 1K DOWNLOADS
- Versions:**
  - Version v1 (Mar 20, 2023) - DOI: 10.5281/zenodo.8153260
  - Version v1 (Mar 20, 2023) - DOI: 10.5281/zenodo.7750935
- Citation:** Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.7750934. This DOI corresponds to the latest version of the dataset.

The screenshot shows a web browser window with the following details:

- Title Bar:** Télécharger un fichier | iLovePDF | Boîte de réception (7735) - kan | Dataset for the genome of med | zenodo.org/records/8153260
- Right Sidebar:**
  - Rights**
  - License:** Creative Commons Attribution 4.0 International
  - Citation:** Qu, Z., Wang, W., & Adelson, D. L. (2023). Dataset for the genome of medicinal plant Sophora flavescens has undergone significant expansion of both transposons and genes (Version v1) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.8153260>
  - Style:** APA
  - Buttons:** Activer Windows, Accept all cookies, Accept only essential cookies, JSON, Export, FRA 17:12, 07/02/2026
- Main Content Area:**
  - Files (4.9 GB):**

Name	Size	Action
Sfla_v1.cdna.gff3	71.2 MB	Download
Sfla_v1.chromosomes.fa	2.1 GB	Download
Sfla_v1.proteins.fa	26.1 MB	Download
Sfla_v1.repeat_EDTA_chromosomes_softMasked.fa	2.1 GB	Download
Sfla_v1.repeat_EDTA.gff3	522.4 MB	Download
Sfla_v1.repeat_EDTA.lib	22.0 MB	Download
Sfla_v1.transcripts.expression_matrix.RSEM TPM.txt	1.6 MB	Preview, Download
  - Information Bar:** This site uses cookies. Find out more on how we use cookies.
  - Bottom Bar:** md5:c7437d3847fe9a28b175a9937b1c2de2f, 14°C, عاصف, 11:10, 07/02/2026

The screenshot shows a web browser window with the following details:

- Title Bar:** Télécharger un fichier | iLovePDF | Boîte de réception (7735) - kan | amelba-ai/Zenodo | Historique des tâches | Nadouch | ChatGPT | zenodo.org/records/8153260
- Right Sidebar:**
  - Details:**
    - DOI: [10.5281/zenodo.8153260](https://doi.org/10.5281/zenodo.8153260)
    - Resource type: Dataset
    - Publisher: Zenodo
  - Rights**
  - Citation:** Sophora flavescens has undergone significant expansion of both transposons and genes
  - Buttons:** Activer Windows, Accept all cookies, Accept only essential cookies, JSON, Export, FRA 11°C, 21:10, 07/02/2026
- Main Content Area:**
  - Table:**

Sfla_0010000002000	0	0	0
Sfla_0010000021000	0	0	0
Sfla_0010000032000	0	0	0
Sfla_0010000023000	0	0	0
Sfla_0010000024000	0	0	0
Sfla_0010000025000	0.19	0.85	0
Sfla_0010000026000	0.11	0	0
Sfla_0010000027000	0.26	1.58	1.7
Sfla_0010000028000	0	0	0
Sfla_0010000029000	0	0	0
Sfla_0010000030000	0	0	0
Sfla_0010000031000	0	0.4	0
Sfla_0010000032000	0	0	0
Sfla_0010000033000	0	0	0
Sfla_0010000034000	0	0	0
Sfla_0010000035000	0	0	0
Sfla_0010000036000	0	0	0
Sfla_0010000037000	0	0	0
  - Files (4.9 GB):**

Name	Size	Action
Sfla_v1.cdna.gff3	71.2 MB	Download
Sfla_v1.chromosomes.fa	2.1 GB	Download
Sfla_v1.proteins.fa	26.1 MB	Download
  - Information Bar:** This site uses cookies. Find out more on how we use cookies.
  - Bottom Bar:** md5:789a0b25cd2e2232011e77e2f, 11°C, عاصف, 21:10, 07/02/2026

Figure : Section Metadata du dataset sur Zenodo, les informations selon le standard Dublin Core.

## ***Conclusion générale :***

libres et les plateformes de science ouverte occupent une place stratégique dans la recherche moderne en sciences de la nature et de la vie. L'étude de Biopython a permis de mettre en évidence la puissance des outils open source dans l'analyse des données biologiques, leur flexibilité, ainsi que leur rôle fondamental dans la reproductibilité et du travail scientifique.

Parallèlement, l'exploration de la plateforme Zenodo illustré l'importance du partage, de l'archivage et de la valorisation des données scientifiques grâce aux principes la science ouverte et à l'attribution des DOI, garantissant traçabilité et la pérennité des travaux de recherche.

Ainsi, l'association entre des outils open source comme Biopython et des plateformes de diffusion comme Zenodo constitue un pilier essentiel pour recherche plus transparente, collaborative et Cette complémentarité favorise non seulement équitable au savoir scientifique, mais aussi continue des pratiques de recherche à l'échelle

