

*Cahier des charges*

*Projet de Programmation*

*L2D1*



*Illustration 1: Image de présentation*

---

---

## Projet L2D -Introduction à la bioinformatique

---

---

*Les informations d'identification du document :*

*Les éléments de vérification du document :*

<b>Référence du document :</b>
<b>Version du document : 2</b>
<b>Date du document : 02/05/2021</b>
<b>Auteurs : Mehdi Hamiche Manal Boutajar Adelin Bodnar</b>

<b>Validé par :</b>
<b>Validé le :</b>
<b>Soumis le :</b>
<b>Type de diffusion :</b>
<b>Confidentialité :</b>

***Les éléments  
d'authentification :***

<b>Maître d'ouvrage :</b>	<b>Chef de projet :</b>
<b>Date / Signature :</b>	<b>Date / Signature :</b>

---

## Projet L2D -Introduction à la bioinformatique

---

### Sommaire

<b><u>1. Introduction</u></b>	<b>5</b>
<b>1.1. Contexte</b>	<b>5</b>
<b>1.2. Historique</b>	<b>6</b>
<b><u>2. Guide de lecture</u></b>	<b>7</b>
<b>2.1. Maîtrise d'œuvre</b>	<b>7</b>
<i>2.1.1. Responsable</i>	<i>7</i>
<i>2.1.2. Personnel administratif</i>	<i>7</i>
<i>2.1.3. Personnel technique</i>	<i>7</i>
<b>2.2. Maîtrise d'ouvrage</b>	<b>8</b>
<i>2.2.1. Responsable</i>	<i>8</i>
<i>2.2.2. Personnel administratif</i>	<i>8</i>
<i>2.2.3. Personnel technique</i>	<i>8</i>
<b><u>3. Description de la demande</u></b>	<b>9</b>
<b>3.2. Produit du projet</b>	<b>10</b>
<b>3.3. Les fonctions du produit</b>	<b>10</b>
<b><u>4. Contraintes</u></b>	<b>12</b>

---

---

## Projet L2D -Introduction à la bioinformatique

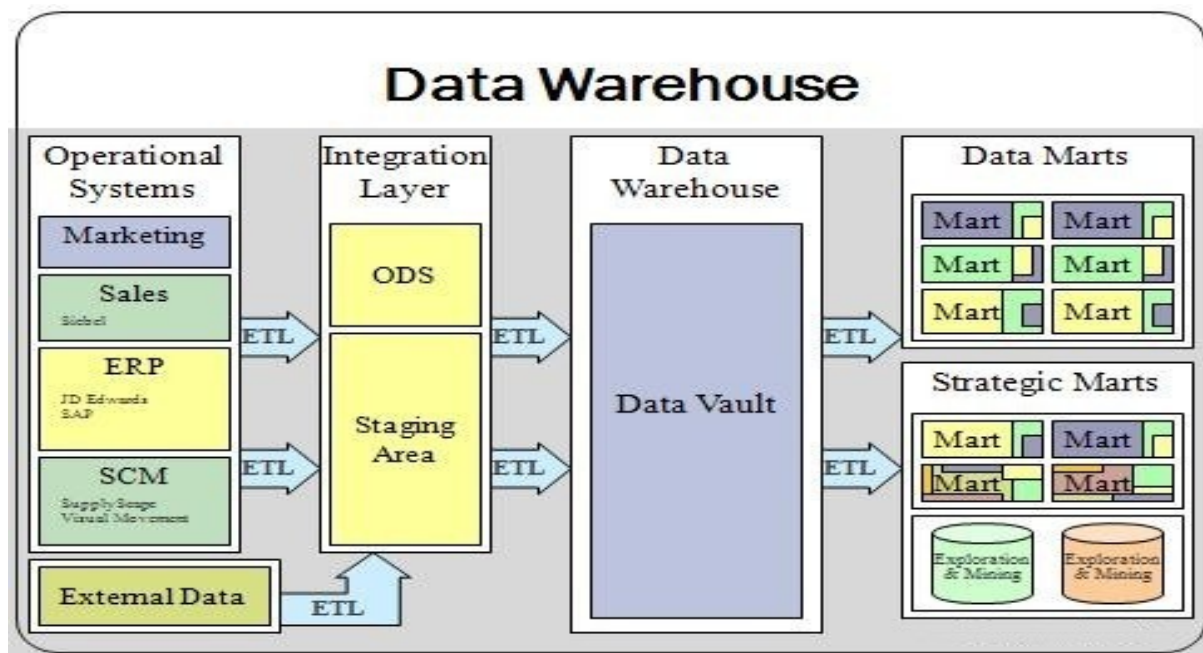
---

<b>4.1. Contraintes de coûts</b>	<b>12</b>
<b>4.2. Contraintes de délais</b>	<b>12</b>
<b>4.3 Contraintes matérielles</b>	<b>12</b>
<b><u>5. Déroulement du projet</u></b>	<b>14</b>
<b>5.1. Planification</b>	<b>14</b>
<b>5.2. Ressources</b>	<b>14</b>
<b><u>6. Annexes</u></b>	<b>15</b>
<b><u>8. Glossaire</u></b>	<b>16</b>
<b><u>9. Références</u></b>	<b>18</b>
<b><u>10. Index</u></b>	<b>19</b>

# 1. Introduction

## 1.1. Contexte

Une discipline qui permet l'analyse de l'information biologique des séquences nucléotidiques et protéiques, voilà ce qu'est la bioinformatique (interdiscipline). Les systèmes biologiques ont des bases / banques de données volumineuses, des algorithmes performants sont nécessaires pour le traitement des informations stockées, ce qui les rend très complexes. Il s'agit en général des traitements qui ressemblent à un traitement traditionnel "data warehouse" avec un calcul plus complexe.



*Illustration 2: Schéma explicatif du « Data Warehouse »*

Nous allons travailler sur les parties traitement et visualisation des données en restant dans le domaine éducatif, en s'appuyant sur l'analyse de la structure primaire (amino acides en lettres) des séquences protéiques (cytochrome c et séquences pour les tests) et des séquences nucléotidiques (gènes) à travers des alignements réalisés.

Ce programme aura une interface graphique pour présenter les résultats des alignements de séquences protéiques et nucléiques.

## 1.2. Historique

Les différents algorithmes permettent la recherche des fonctions biologiques, de la phylogénie moléculaire, du séquençage de génome etc.

Exemples d' algorithmes : Smith & Waterman, Needleman & Wunsch, Réseaux bayésiens.

La bioinformatique propose des méthodes et des outils pour le recueil des données, leurs traitements, leurs distributions et leurs visualisations.

## 2. Guide de lecture

---

### 2.1. Maîtrise d'œuvre

La maîtrise d'œuvre présente l'équipe du développement chargé du bon suivi du cahier des charges et des besoins dont le maître d'ouvrage fait commande.

Elle représente l'équipe du développement :

- Adelin Bodnar
- Manal Boutajar
- Mehdi Hamiche

Cette équipe veillera au bon suivi du cahier des charges représentant les besoins des enseignants encadrants.

#### *2.1.1. Responsable*

Il est conseillé pour le responsable de la maîtrise d'œuvre de lire le document dans sa totalité afin de prendre conscience de l'ensemble des éléments.

#### *2.1.2. Personnel administratif*

Il est conseillé pour le personnel administratif de lire la présentation du produit, les contraintes, le planning et les livrables.

#### *2.1.3. Personnel technique*

Il est conseillé pour le personnel technique de prendre compte les parties sur les concepts de base, la description du produit, l'aspect visuel et navigation ainsi que les contraintes.

## 2.2. Maîtrise d'ouvrage

La maîtrise d'ouvrage représente dans notre cas le client du projet, c'est-à-dire les personnes dont les besoins permettent la conception du projet.

La maîtrise d'ouvrage est assistée par l'équipe de la maîtrise d'œuvre et donc ce rôle sera assuré par les enseignants encadrants Dragutin Jastrebic et Koviljka Lukic Jastrebic.

### *2.2.1. Responsable*

Il est conseillé pour le responsable de la maîtrise d'ouvrage de lire le document dans toute sa totalité afin de prendre conscience de l'ensemble des documents.

### *2.2.2. Personnel administratif*

Il est conseillé pour le personnel administratif de lire la présentation du produit, les contraintes, le planning et les livrables.

### *2.2.3. Personnel technique*

Il est conseillé pour le personnel technique de prendre compte les parties sur les concepts de base, la description du produit, l'aspect visuel et navigation ainsi que les contraintes.



## 3. Description de la demande

---

- Un programme capable d'analyser et d'aligner des séquences protéiques,
- Accéder aux bases biologiques,
- Manipuler les listes, fonctions, objets, matrices,
- 3 approches : manuel, programmable et production,
- Les protéines sont stockées dans des bases de données,
- Les types d'alignements : global, local, multiple,
- Affichage des résultats.
- Découvrir la diversité des formats / nature des données biologiques,
- Base de données biologiques pour extraire les données d'intérêt / outils et programmes bioinformatiques / langages informatiques,
- Base de données pour stocker les données / système de fichiers / protocoles d'échange de données.

### Local Alignment

Target Sequence  
5' ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA 3'  
Query Sequence 5' TACTCACGGATGAGGTACTTTAGAGGC 3'

### Global Alignment

Target Sequence  
5' ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA 3'  
Query Sequence 5' ACTACTAGATT---ACGGATC--GTACTTTAGAGGCTAGCAACCA 3'

**Illustration 3: Différence entre alignement global / local**

Pour ce fait, on aura besoin des technologies suivantes :

- ❖ HTML : Ce langage est utilisé pour définir la structure des pages,
- ❖ CSS : Ce langage sert à styliser l'ensemble des pages,
- ❖ Java : Ce langage va nous servir pour créer notre programme.

## 3.2. Produit du projet

Le produit final est un programme permettant de traiter les données stockées et de présenter les résultats des alignements de séquences protéiques et nucléotidiques. Le but est de produire un programme en développant sur JAVA.

- La nature de séquence à comparer : séquences nucléotidiques et séquences protéiques,
- Séquences à aligner : Cytochrome c, Insuline et Séquences nucléiques
- Exemple de l'ADN : ACTG / CTTG (séquences comparée),
- Pour l'ARN : ACUG,
- Séquences pour les test : Cytochrome c, Insuline et Séquences nucléiques



*Illustration 4:*  
*Image de*  
*Cytochrome*

Ce projet devra donc s'adapter aux besoins de l'utilisateur et à son environnement, par étudier leurs instructions et les prendre en considération à chaque avancement.

## 3.3. Les fonctions du produit

L'alignement des séquences protéiques ou nucléiques peut être :

- A) Global - on aligne la totalité des 2 séquences,
- B) Multiple - on aligne plusieurs séquences pour faire le logo.

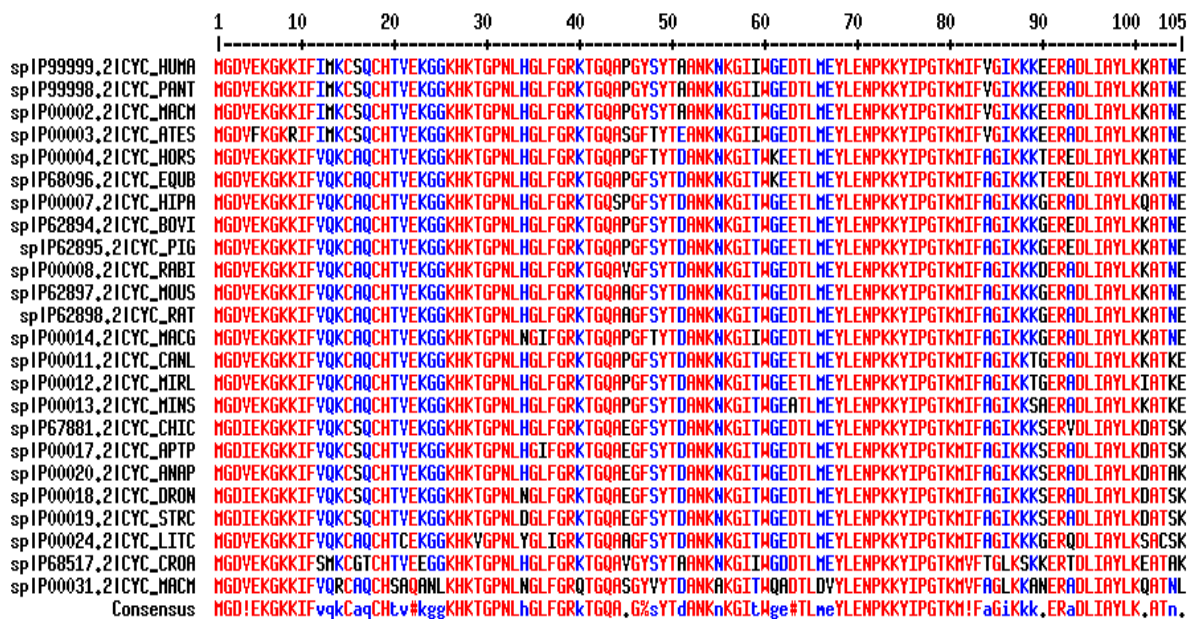


Illustration 5: Exemple d'alignement multiple

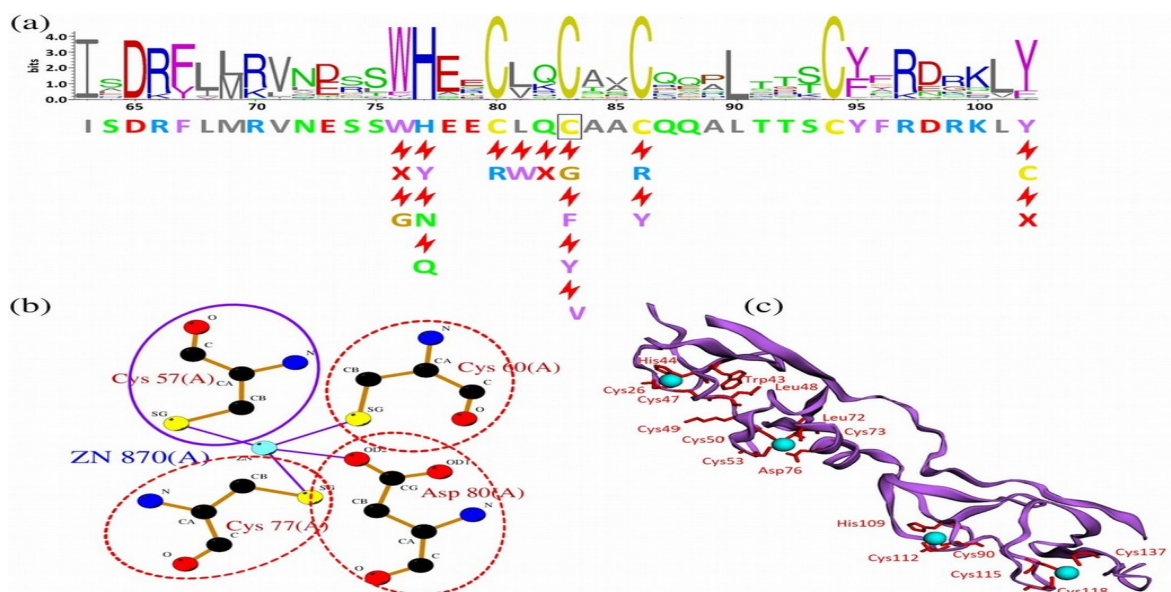


Illustration 6: Exemple de logo

## 4. Contraintes

---

### 4.1. Contraintes de coûts

Ce projet a un but éducatif, son budget est nul et les membres du groupe ne seront pas rémunérés.

### 4.2. Contraintes de délais

- *Le cahier des charges* (à rendre semaine 3);
- *Le cahier de recettes* (à rendre semaine 4);
- *La conception générale*;
- *La conception détaillée* (à rendre semaine 5);
- *Le manuel d'utilisation* (à rendre semaine 11);
- *Le manuel d'installation* (à rendre semaine 11);
- *Le plan de tests* (à rendre semaine 11);
- *La documentation interne du code* (à rendre semaine 11);
- *Le code sources du programme* (à rendre semaine 11);
- *Le rapport du projet* (avant la soutenance);
- *Le résumé en français et en anglais* (avant la soutenance);
- *Les diapositives sonorisées* (avant la soutenance).
  - Date de fermeture de la forge : **16/05/2021**

### 4.3 Contraintes matérielles

- ❖ Wamp dernière version (Wamp inclut PHP, Apache, les bases MySQL et MariaDB);
- ❖ Python 3x;

---

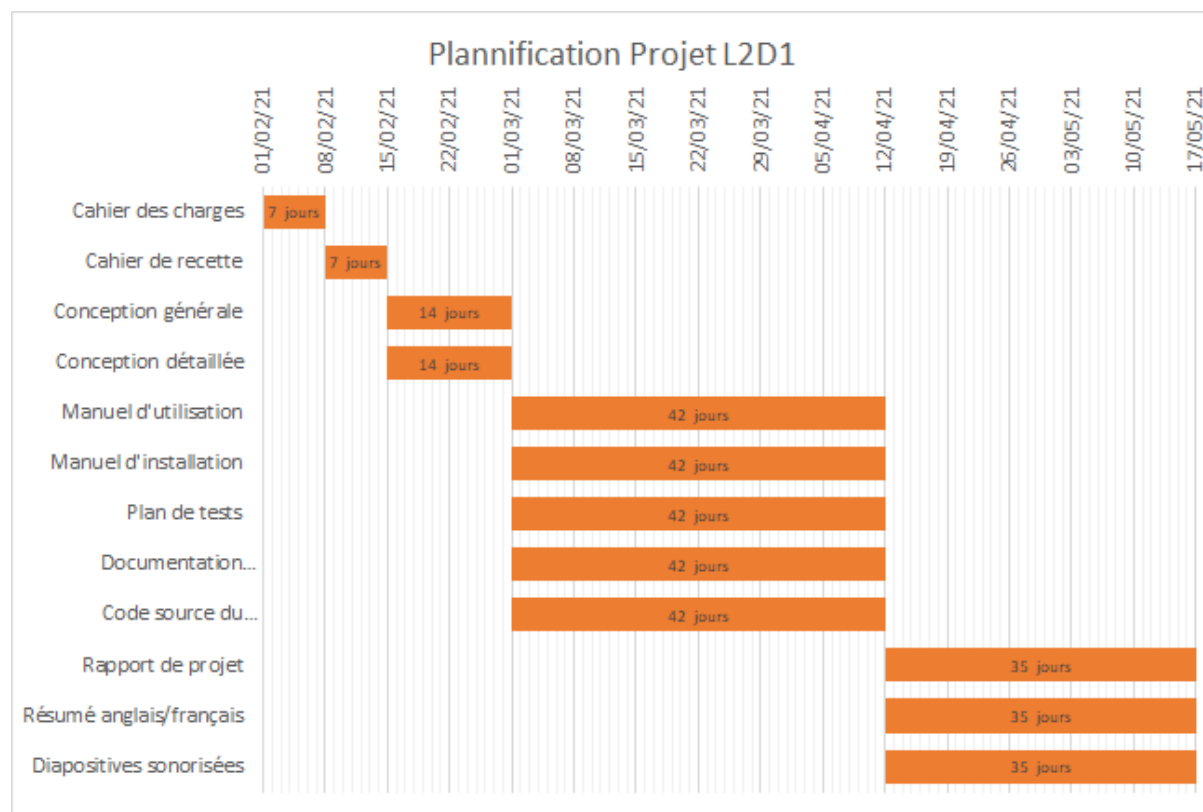
## Projet L2D - Introduction à la bioinformatique

---

- ❖ Java Eclipse;
- ❖ R et Rstudio.

## 5. Déroulement du projet

### 5.1. Planification



*Illustration 7: Diagramme de Gantt*

### 5.2. Ressources

L'aide et les conseils des encadrants. Quelques sources sur internet peuvent être utilisées aussi.

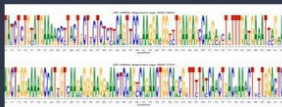
# 6. Annexes

## Maquette Figma

### Introduction à la bioinformatique

#### Maquette du produit

#### Description du projet



#### Groupe L2D1

- ❖ Programme qui analyse et aligne des séquences protéiques,
- ❖ Accéder aux bases biologiques,
- ❖ Manipuler les listes, fonctions, objets, matrices,
- ❖ 3 approches : manuel, programmable et production,
- ❖ Les protéines sont stockées dans des bases de données,
- ❖ Les types d'alignements : global, local, multiple,
- ❖ Affichage des résultats.

#### Technologie utilisées :

- **HTML** : Ce langage sera utilisé pour définir la structure des pages,
- **CSS** : Ce langage servira à styliser l'ensemble des pages,
- **PHP** : Ce langage permettra la communication entre le client et le serveur,
- **Javascript** : Afin d'animer nos pages web.

#### Objectif

- ❖ Travailler sur les parties traitement et visualisation des données - domaine éducatif,
- ❖ Analyse de la structure primaire (amino acides en lettres) des séquences protéiques (cytochrome c et séquences pour les tests) et des séquences nucléiques (gènes) à travers des alignements réalisés,
- ❖ Présenter les résultats des alignements de séquences protéiques et nucléique avec interface graphique.

#### Différents délais



- **Le cahier des charges** (à rendre semaine 3);
- **Le cahier de recette** (à rendre semaine 4);
- **La conception générale**;
- **La conception détaillée** (à rendre semaine 5);
- **Le manuel d'utilisation** (à rendre semaine 11);
- **Le manuel d'installation** (à rendre semaine 11);
- **Le plan de tests** (à rendre semaine 11);
- **La documentation interne du code** (à rendre semaine 11);
- **Le code sources du programme** (à rendre semaine 11);
- **Le rapport du projet** (avant la soutenance);
- **Le résumé en français et en anglais** (avant la soutenance);
- **Les diapositives sonorisées** (avant la soutenance).

## 8. Glossaire

---

- ★ **BIO-INFORMATIQUE** : discipline permettant d'analyser de l'information biologique qu'il y a dans les séquences nucléotidiques et protéiques
- ★ **PROGRAMMATION ORIENTÉE OBJET** : met en œuvre différents objets (instances de classes). Chaque objet associe des données et des méthodes (fonctions) agissant exclusivement sur les données
- ★ **ALGORITHME NEEDLEMAN - WUNSCH** : effectue un alignement global maximal de deux chaînes de caractères
- ★ **ALIGNEMENT GLOBAL** : recouvre les séquences alignées (par exemple 2 séquences) sur l'ensemble de leur longueur
- ★ **ALIGNEMENT MULTIPLE** : aligner un ensemble de séquences homologues, comme des séquences protéiques ou nucléiques qui assure des fonctions similaires dans différentes espèces vivantes
- ★ **WEBLOGO** : application web conçue pour rendre la génération de logos de séquence aussi simple que possible
- ★ **INTERFACE GRAPHIQUE** : relie les 3 modules et les présente de manière simple et efficace
- ★ **GAP, MATCH, MISMATCH** : valeurs pour alignement global
- ★ **FASTA** : format de fichier texte dans lesquels les séquences sont affichées par une suite de lettres
- ★ **NCBI** : National Centre for Biotechnology Information
- ★ **PROTEIN** : Base de données pour la récupération des séquences protéiques en format .fasta
- ★ **GENE** : Base de données pour récupération des séquences nucléiques en format .fasta



*Illustration 8:*  
**Logo NCBI**



---

## Projet L2D - Introduction à la bioinformatique

---

- ★ **NUCLEOTIDE** : Base de données pour la récupération des séquences nucléiques en format .fasta
- ★ **UNIPROT** : Base de données pour la récupération des séquences protéiques en format .fasta
- ★ **DDBJ** : Base de données pour la récupération des séquences nucléiques en format .fasta
  
- ★ **DNA** : Data Bank Japan
- ★ **EMBL** : European Molecular Biology Laboratory
  
- ★ **EMBOSS** : outil bioinformatique pour l'alignement global - NEEDLE
- ★ **BLAST** : outil bioinformatique pour l'alignement local
- ★ **MULTALIN** : outil bioinformatique pour l'alignement multiple



*Illustration 9: Logo EMBL*



*Illustration 10:  
Logo DNA*



*Illustration 11: Logo  
DDBJ*

## 9. Références

---

[Alignement de séquences](#)

[BNLEARN](#)

[DDBJ](#)

[National Centre for Biotechnology Information](#)

[Réseaux Bayésiens](#)

[UniProt](#)

## 10. Index

---

### Index des figures

<i>Illustration 1: Image de présentation</i>	<i>1</i>
<i>Illustration 2: Schéma explicatif du « Data Warehouse »</i>	<i>5</i>
<i>Illustration 3: Différence entre alignement global / local</i>	<i>9</i>
<i>Illustration 4: Image de Cytochrome</i>	<i>10</i>
<i>Illustration 5: Exemple d'alignement multiple</i>	<i>11</i>
<i>Illustration 6: Exemple de logo</i>	<i>11</i>
<i>Illustration 7: Diagramme de Gantt</i>	<i>14</i>
<i>Illustration 8: Logo NCBI</i>	<i>16</i>
<i>Illustration 9: Logo EMBL</i>	<i>17</i>
<i>Illustration 10: Logo DNA</i>	<i>17</i>
<i>Illustration 11: Logo DDBJ</i>	<i>17</i>

### Index lexical

ALGORITHME NEEDLEMAN - WUNSCH :	16
Alignement de séquences	18
alignement global	16 sv
ALIGNEMENT GLOBAL	16

---



---

## Projet L2D - Introduction à la bioinformatique

---



---

alignement multiple	17
ALIGNEMENT MULTIPLE	16
alignements	5 sv, 9 sv
application	16
Base de données	9, 16 sv
BIO-INFORMATIQUE	16
bioinformatique	5 sv, 9, 17
BNLEARN	18
CSS	9
data warehouse	5
DDBJ	18
fasta	16 sv
FASTA	16
GAP	16
global	9, 16 sv
Global	10
GLOBAL	16
HTML	9

---



---

## Projet L2D - Introduction à la bioinformatique

---



---

interdiscipline	5
interface graphique	6
Interface graphique	16
INTERFACE GRAPHIQUE	16
Java	9, 13
JAVA	10
logo	10, 16
LOGO	16
MATCH	16
MISMATCH	16
MULTALIN :	17
multiple	9, 17
Multiple	10
MULTIPLE	16
National Centre for Biotechnology Information	18
NEEDLEMAN - WUNSCH	16
Needleman & Wunsch	6
nucléiques	17

---



---

## Projet L2D - Introduction à la bioinformatique

---



---

nucléotidiques	16
projet	8, 10, 12, 14
Projet	1
protéiques	5 sv, 9 sv, 16 sv
Python	13
Réseaux bayésiens	6
Réseaux Bayésiens	18
Rstudio	13
séquences nucléotidiques	5, 10, 16
séquences protéiques	16 sv
Smith & Waterman	6
systèmes biologiques	5
UniProt	18
Wamp	12
WEBLOGO	16
	17