

# GenAnnot.

UE Programmation WEB- Rendu Livrable 2

**Encadrement :** Sarah Cohen-Boulakia & Olivier Lespinet  
*Ilhem Riahi, Lorraine Weng, Arnaud Maupas, Sanâ Cherifi*



## Table des Matières

<b>Documentation Utilisateur</b>	<b>3</b>
Introduction	3
Base de données	3
Page d'accueil	3
Page d'inscription	4
Page de connexion	5
Page lecteur	6
Page Annotateur	8
Page Validateur	8
Page Administrateur	9
<b>Documentation Technique</b>	<b>10</b>
Logiciels et outils utilisés	10
Chargement des données	11
<b>Retour Expérience</b>	<b>11</b>
<b>Sources</b>	<b>12</b>

# Documentation Utilisateur

## I. Introduction

L'objectif principal de ce projet est de créer une application web pour l'annotation fonctionnelle de génomes et séquences de gènes et protéines. Nous avons ainsi créé le site web "GenAnnot", afin de permettre à des biologistes et bio-informaticiens d'annoter les fonctions des séquences génomiques et protéiques, issues de génomes procaryotes et de les étudier. Pour accéder à ces outils et parcourir le site, il est tout d'abord nécessaire de s'inscrire, puis de se connecter (accès restreint). Nous avons voulu mettre au point un site ergonomique et dynamique pour l'utilisateur afin de rendre l'utilisation de l'application la plus simple, guidée et intuitive possible. Nous présenterons ici une description détaillée de cet outil, et nous montrerons ainsi toutes les étapes et recommandations nécessaires à sa bonne utilisation.

## II. Base de données

A ce jour, trois génomes d'*Escherichia coli* sont stockés dans la base de données avec leurs annotations (k12, CFT073 et o157h7) et un nouveau génome (new\_coli) est en attente d'annotation. Les génomes et leurs CDS (Coding DNA Sequence, séquence codante) et peptides associés des *Escherichia coli* sont considérés comme "annotés et validés" dans notre base de données.

### 1. Page d'accueil "Home\_page.php"

La première page sur laquelle se rend un visiteur du site est la page d'accueil, elle affiche le nom du site "GenAnnot", ainsi qu'un message d'accueil indiquant une description du site web et les fonctionnalités qu'il comporte.

**Welcome to GenAnnot !**

This web application was created by four students in Computational Biology (AMI2B), in Paris-Saclay University. GenAnnot is full of information about bacterial genomes, genes and proteins.

The aim of the website is to make functional annotations on genes and proteins. The new annotations will be validated or rejected before submitting them to the database. The database is written in PostgreSQL. As an user you have the right to choose your role as a reader, validator or annotator:

- If you sign up as a reader, you can only search, read information about bacterial genomes, genes and protein. You can also check the annotation in the process of validation and download the database data.
- However, if you sign up as an annotator or validator, you will be able to perform annotation and functional analysis on bacterial genome.
- Otherwise, if you choose a validator role you have the permission to validate or reject annotations. You are totally free to choose the role you want.
- There is also the admin page for the GenAnnot Admin, where the admin can consult information about registered users and delete unwanted users. Nevertheless, the admin role can't be chosen in the sign in form.

**Important information :** If you already have an account, you can't re-register on the website. Only existing users in the website can check and search information in the GenAnnot database.

**Figure 1 :** Page d'accueil du site GenAnnot

En haut se trouve un menu de navigation avec différentes fonctionnalités. La couleur du nom des menus change lorsque la souris passe dessus. Nous y trouvons le bouton "Sign Up" et le bouton "Sign in" qui sont apposés dans le coin supérieur droit. Ces deux boutons permettent respectivement à tout nouvel utilisateur de créer un compte et de se connecter dans le cas d'un utilisateur déjà enregistré dans la base de données.

## 2. Page d'inscription "Signup.php"

Dans l'espace dédié à l'inscription, l'utilisateur se doit d'entrer ses informations personnelles, à savoir son nom et prénom, son email, son numéro de téléphone et son mot de passe (un champ "confirmation de mot de passe" est également prévu). Enfin il peut choisir le rôle qu'il souhaite endosser au sein du site par le biais d'une liste déroulante avec les 3 rôles "lecteur", "annotateur", "validateur". Une fois les champs remplis, les informations sont sauvegardées dans la base en cliquant sur le bouton "Register". Un message apparaît alors en indiquant "Your registration has been successfully registered". Autrement, si le format de l'email est incorrect ou de mot de passe , un message d'erreur apparaît.

Concernant les différents rôles possibles, la page lecteur est commune à tous les utilisateurs en dépit de leur rôle au sein du site web. Elle permet d'accéder notamment aux fonctionnalités de recherche d'informations sur les séquences, la visualisation du génome, des annotations, etc. En revanche, l'annotation de séquences (page Annotator) ne sera accessible qu'aux annotateurs et validateurs. De même, la page validator permettant d'assigner et contrôler les annotations (validation ou rejet) est dédiée uniquement aux validateurs. Enfin l'administrateur du site gère son fonctionnement dans sa globalité en ayant accès à toutes ces pages, en plus de son espace dédié. Cependant, il n'est pas possible de choisir ce rôle lors de l'inscription.

---

**GenAnnot.**

Home Sign Up Sign In

The form fields are as follows:

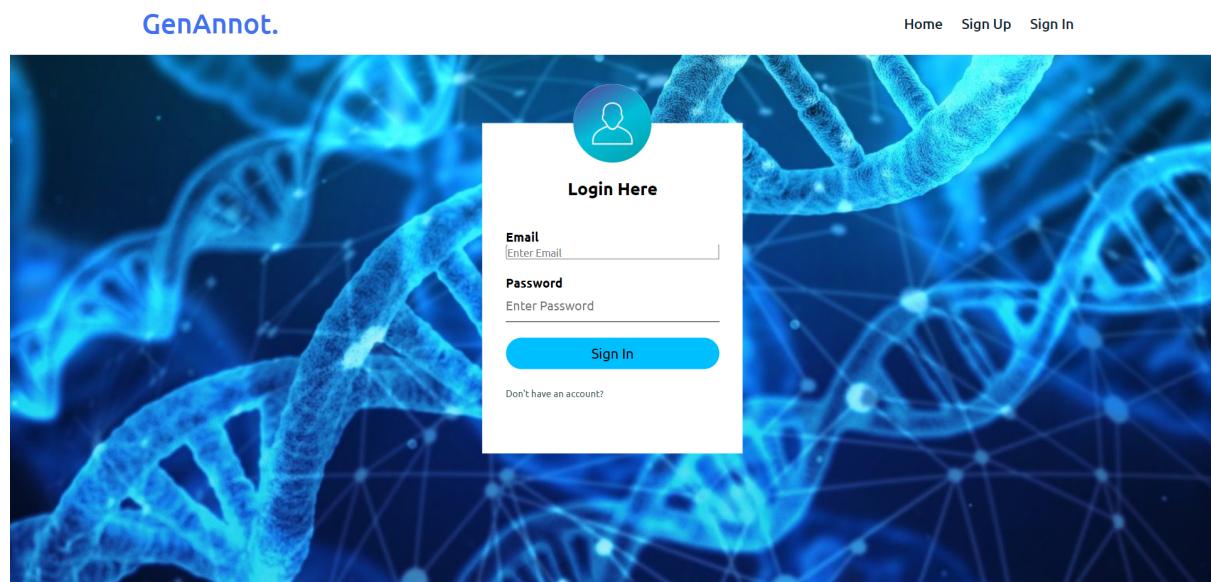
- Name: Enter your name
- Email: Enter your Email
- Password: Enter your password
- Family name: Enter your Family name
- Phone Number: Enter your Number
- Confirm your password: Enter your password
- Choose your role: Reader

**Register**

**Figure 2 :** Page d'inscription

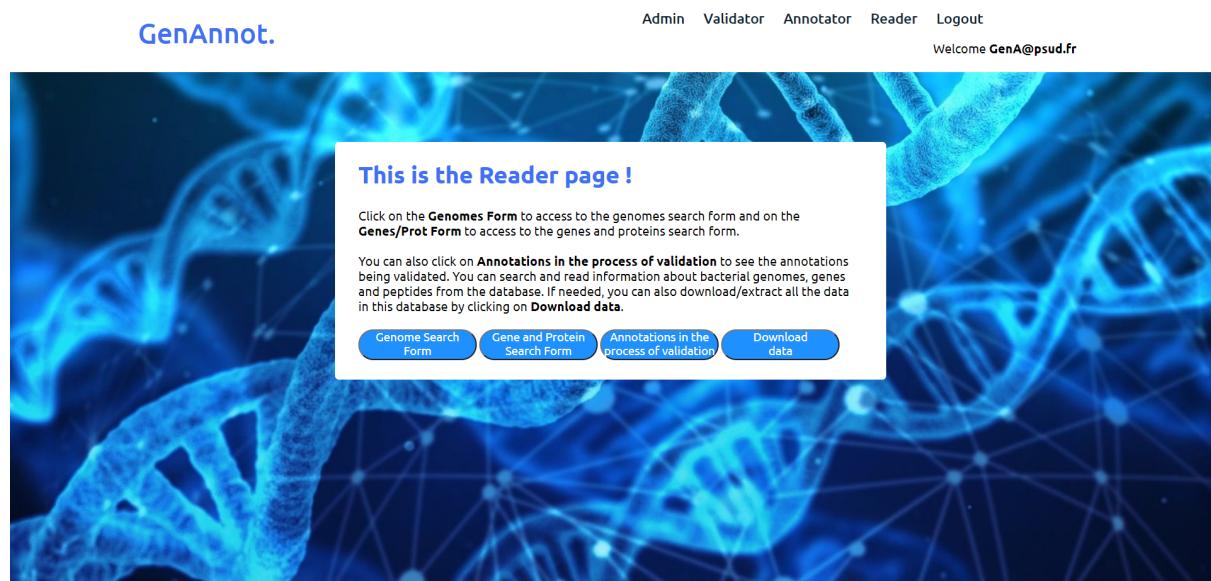
### 3. Page de connexion “Signin.php”

Une fois inscrit, l'utilisateur pourra se connecter à partir de son adresse e-mail et de son mot de passe. Nous avons également ajouté un lien cliquable “Don't have an account”, qui redirige vers la page d'inscription dans le cas où l'utilisateur ne possèderait pas de compte. Également, si les données d'inscription ne sont pas reconnues (pas de match avec la base de données) alors l'utilisateur recevra un message d'erreur “Username/password is incorrect”. Une fois connecté, l'utilisateur est renvoyé vers l'espace qui lui est dédié selon son rôle précédemment choisi. La barre de navigation en haut de la page est différente pour chaque rôle et permet d'accéder à un espace personnalisé en fonction du rôle de l'utilisateur. Étant donné cet accès restreint, le rôle détermine les pages visibles et les actions possibles pour chaque utilisateur. La barre de menu au complet présente les boutons “Admin” , “Validator”, “Annotator”, “Reader” et “Logout”. Le “Logout”, comme son nom l'indique, permet à l'utilisateur de se déconnecter.

**Figure 3 :** Page de connexion

#### 4. Page lecteur “reader\_Menu.php”

Comme indiqué précédemment, n'importe quel utilisateur inscrit a accès à la page lecteur et est donc en mesure de lancer une recherche de génome ou de séquence (nucléique ou peptidique). Ainsi, à l'entrée de la page Reader, un message de bienvenue s'affiche sur lequel sont apposés quatre boutons permettant d'accéder à la page “Genome Form” (recherche de génome), “Gene/Prot Form” (recherche de séquences géniques et protéiques), “Annotations in the process of validation” et “Download data”. Les deux derniers boutons permettent respectivement de consulter les annotations en cours de validation (“annotat\_in\_progress.php”) et de télécharger toutes les données de la base de données, en checkant les attributs que l'utilisateur veut extraire (“extraction.php”).

**Figure 4 :** Page d'accueil Lecteur

a) Formulaires Genome “Form\_genome.php” et Gene/prot “Form\_cds.php”

Sur la page “GenomeForm”, l’utilisateur retrouve un formulaire à compléter afin de lancer sa recherche sur un génome en particulier. Plusieurs champs sont ainsi prévus à cet effet : le numéro d’accession, le nom de l’espèce, le nom de la souche, longueur de la séquence et la séquence du génome complet.

La page Gene/ProtForm est basée sur le même principe et rassemble tous les attributs relatifs à la séquence et sa fonction (ID de la séquence, numéro d’accession du génome, type ADN, longueur et position de début et de fin du CDS, brin direct ou indirect, taille du CDS, taille du peptide, ID du gène, biotype du transcrit et du gène, symbole du gène, description du CDS). Dans les deux cas formulaires, plusieurs champs peuvent être combinés au sein d’une même requête et la recherche considère tous les CDS et génomes disponibles. Il est également possible de faire une recherche par le nom de l’espèce “Escherichia\_coli” et “ATGC” pour obtenir les séquences comportant au moins le motif ATGC.

b) Liste des résultats de la recherche “Search\_genome.php” et “Search\_cds.php”

Une fois la recherche lancée (via le bouton search), l’utilisateur est renvoyé vers une page recensant une liste de tous les génomes ou CDS et peptides répondant à sa recherche. Les ID de séquences et les numéros de d’accession de génomes ont ainsi été rendus cliquables afin de permettre aux utilisateurs d’accéder aux informations propres et plus détaillées à un génome ou CDS (et donc protéine) donné.

c) Les fiches résultats des genome, CDS/protéines “Results\_genome2.php” et “Results\_cds.php”

Sur la page de résultat du génome, nous affichons tout d’abord les mêmes informations que celles présentes dans le formulaire en faisant apparaître le numéro d’accession, l’espèce, la souche et la taille de la séquence. De plus, nous pouvons également retrouver une fenêtre graphique que l’on peut faire défiler affichant la séquence nucléotidique associée. Il faut pour cela définir les positions de départ et fin de la séquence à afficher (0 et 5000 par défaut). Il est recommandé de ne pas dépasser une longueur de 5000 paires de bases pour garantir un résultat optimal. Nous pouvons repérer dans cette visualisation du génome entier quels loci correspondent à quels CDS (par l’identifiant de séquence situé en dessous de la portion du gène d’intérêt) et ainsi accéder à la fiche résultat du CDS en question en cliquant sur l’ID de séquence indiqué si il est annoté.

La page de résultats des gènes et protéines permet d’afficher un CDS particulier accompagné de la séquence protéique qui lui est associée. On retrouve à nouveau les mêmes informations que dans le formulaire de recherche. En complément des informations disponibles sur GenAnnot, il est également possible de consulter les fiches de banques de données externes comme NCBI, Uniprot et ensembl bacteria via l’ID de la séquence, depuis les boutons respectifs, en bas de la page des résultats. Une autre fonctionnalité possible est

l'utilisation de blastn et blastp pour lancer l'alignement multiple de la séquence nucléotidique ou protéique d'intérêt (en appelant les scripts "blast\_cds.php" et "blastprot.php").

## 5. Page Annotateur “Annot\_Menu.php”

La possibilité d'annoter est uniquement accordée aux annotateurs et aux validateurs. Les annotateurs et les validateurs peuvent annoter des séquences qui leur sont attribuées par l'administrateur ou un validateur. Ainsi en cliquant sur le menu Annotator, cela amène à la page d'accueil des annotateurs avec un message et un bouton "Choose a sequence to annotate", qui permet d'accéder à une page recensant les différentes séquences lui ayant été attribuées. Celui-ci peut donc choisir une séquence en cliquant sur l'ID de la séquence (page "annot\_seq.php"). Il sera alors en mesure de choisir une séquence et compléter les champs d'annotations correspondants. Dans le cas où aucune séquence ne lui a encore été attribuée, un message "Unfortunately there is no sequence to annotate" est affiché sur la page.

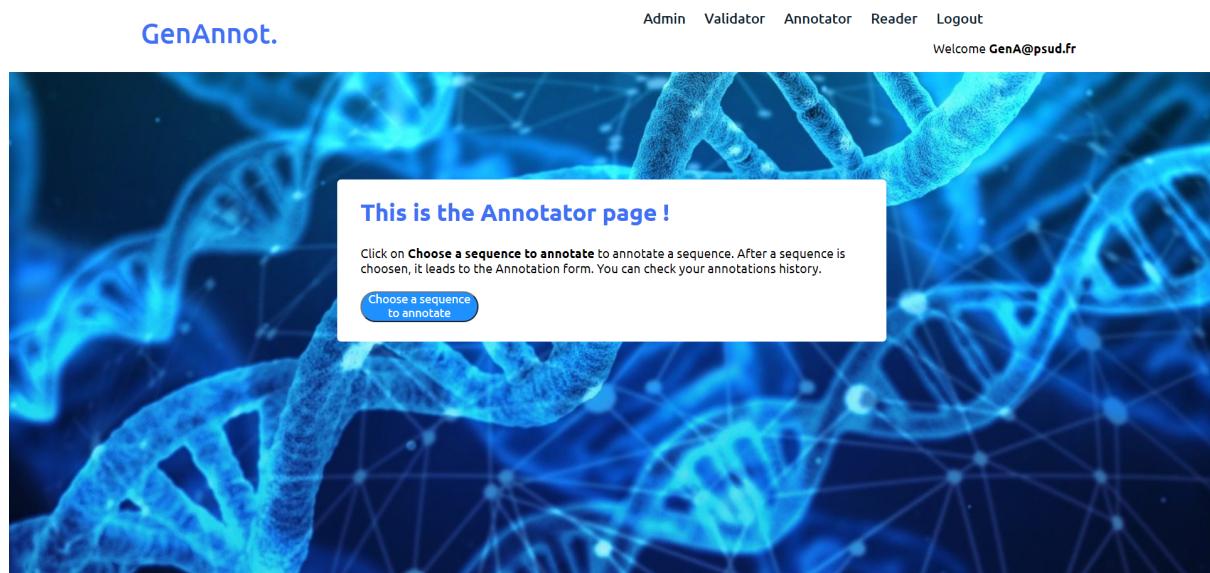


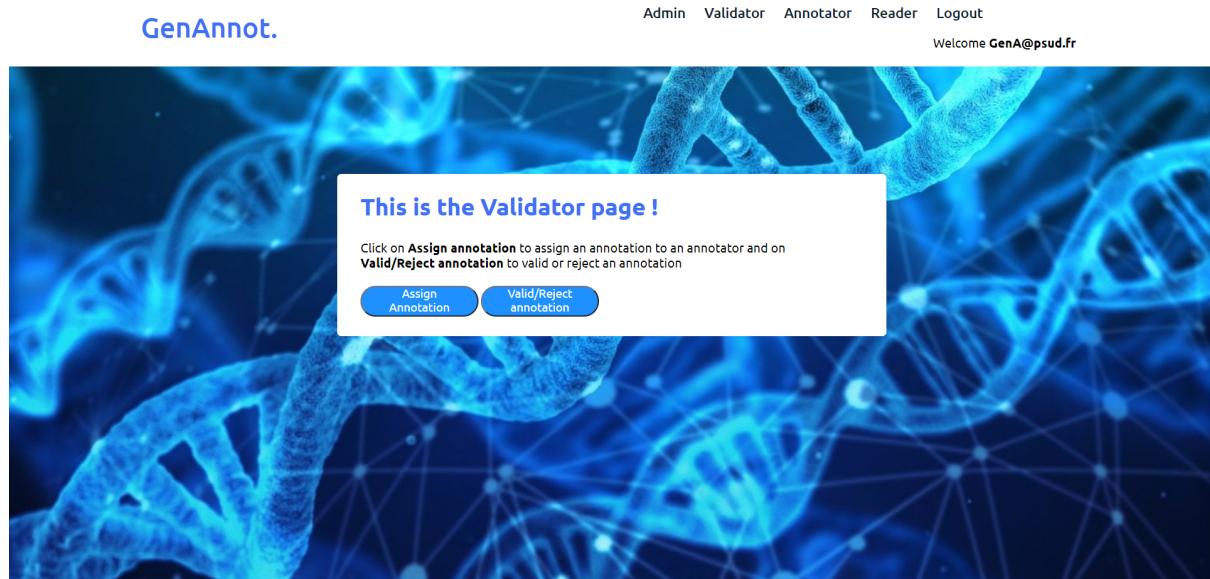
Figure 5 : Page Annotateur

En cliquant sur le bouton de l'ID séquence, on obtient le formulaire pour ajouter des annotations fonctionnelles ("annot\_design2.php"). L'annotateur peut également consulter l'historique de ses annotations ("historique.php").

## 6. Page Validateur “Validator\_Menu.php”

La page validator permettra aux validateurs tout d'abord d'assigner des annotations aux annotateurs ("assign\_annot.php"). Il sélectionne pour cela l'email de l'annotateur choisi dans une liste déroulante et valide sa sélection avec le bouton "assign". Le validateur peut également visualiser les annotations ayant été faites, afin de les valider ou non.

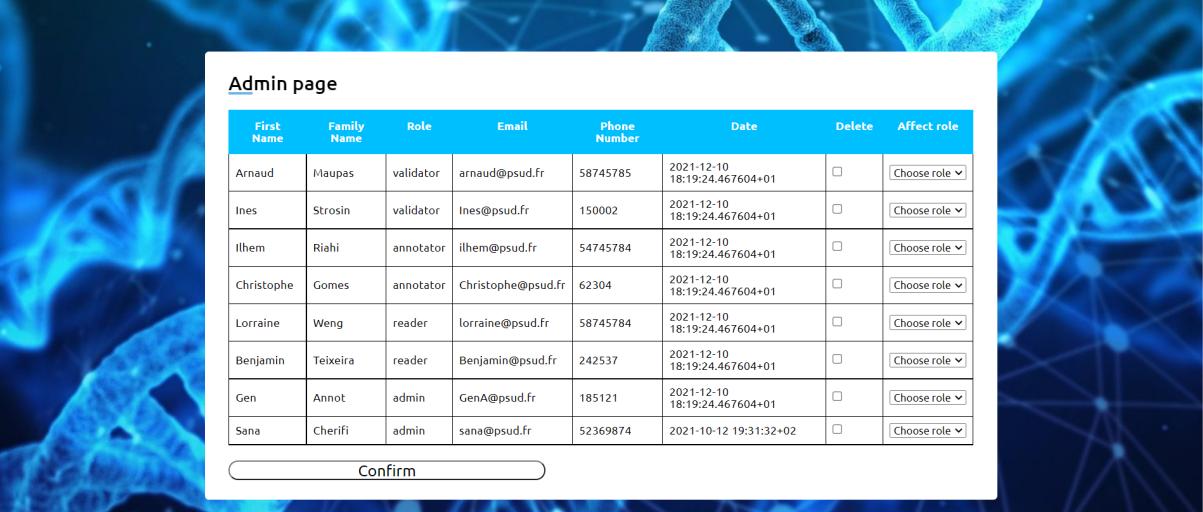
("validation\_space.php") avec un commentaire. Les séquences refusées seront ensuite réaffectées à un autre annotateur dans la page assign\_annot.php.



**Figure 6:** Page Validateur

## 7. Page Administrateur “adminpage2.php” (et adminmodif2.php)

Enfin l'administrateur a accès à une page Admin dans laquelle il peut visualiser tous les profils des utilisateurs inscrits sur le site. La page administrateur n'est accessible que pour les emails identifiés préalablement dans la base de données comme étant des administrateurs. En effet, il n'est pas possible de choisir le rôle “administrateur” dans la page d'inscription ou de connexion. Un administrateur peut consulter les données des utilisateurs rentrées dans le formulaire d'inscription. Mais il peut également consulter la date et heure de dernière connexion (ou d'inscription, s'il ne s'est jamais connecté depuis son inscription), supprimer un utilisateur en cochant la case “delete” du checkbox et changer le rôle d'un utilisateur. Un administrateur possède par ailleurs les mêmes permissions que tous les autres types de rôles.



The screenshot shows a web application interface titled "Admin page". At the top right, there are links for "Admin", "Validator", "Annotator", "Reader", "Logout", and a welcome message "Welcome sana@psud.fr". The main content is a table listing users with columns: First Name, Family Name, Role, Email, Phone Number, Date, Delete, and Affect role. Each row contains a "Choose role" dropdown menu. A "Confirm" button is at the bottom of the table.

First Name	Family Name	Role	Email	Phone Number	Date	Delete	Affect role
Arnaud	Maupas	validator	arnaud@psud.fr	58745785	2021-12-10 18:19:24.467604+01	<input type="checkbox"/>	<input type="button" value="Choose role ▾"/>
Ines	Strosin	validator	Ines@psud.fr	150002	2021-12-10 18:19:24.467604+01	<input type="checkbox"/>	<input type="button" value="Choose role ▾"/>
Ilhem	Riahi	annotator	ilhem@psud.fr	54745784	2021-12-10 18:19:24.467604+01	<input type="checkbox"/>	<input type="button" value="Choose role ▾"/>
Christophe	Gomes	annotator	Christophe@psud.fr	62304	2021-12-10 18:19:24.467604+01	<input type="checkbox"/>	<input type="button" value="Choose role ▾"/>
Lorraine	Weng	reader	lorraine@psud.fr	58745784	2021-12-10 18:19:24.467604+01	<input type="checkbox"/>	<input type="button" value="Choose role ▾"/>
Benjamin	Teixeira	reader	Benjamin@psud.fr	242537	2021-12-10 18:19:24.467604+01	<input type="checkbox"/>	<input type="button" value="Choose role ▾"/>
Gen	Annot	admin	GenA@psud.fr	185121	2021-12-10 18:19:24.467604+01	<input type="checkbox"/>	<input type="button" value="Choose role ▾"/>
Sana	Cherifi	admin	sana@psud.fr	52369874	2021-10-12 19:31:32+02	<input type="checkbox"/>	<input type="button" value="Choose role ▾"/>

**Figure 7:** Page Administrateur

## Documentation Technique

### I. Logiciels et outils utilisés

Pour mener à bien ce projet de création de site web, nous nous sommes servis de plusieurs outils différents.

Dans un premier temps, nous nous sommes servis du logiciel PostgreSQL version 14.0 pour la base de données et des langages HTML et CSS ainsi que PHP. Nous avons également utilisé Python version 3.8 (avec Spyder et PyCharm) pour parser les fichiers fasta à insérer dans notre base de données. Pour les environnements PHP, nous avons tout d'abord travaillé sur l'IDE (Environnement de développement) Visual Studio Code, puis sur PHPstorm version 11.0.13 car ce dernier présente des fonctionnalités plus avancées comme notamment la possibilité de visualiser les données de la base de données PostgreSQL directement dans cet IDE. Certains membres du groupe ont également choisi de travailler avec l'interface pgAdmin4 pour la gestion de la base de données. Afin de coder les pages du site à proprement parler nous avons utilisé divers langages de programmation. Tout d'abord, nous avons réalisé des maquettes des pages Web en HTML, nous avons également créé pour chaque page un fichier .css permettant d'organiser le style, l'aspect et l'affichage des différentes pages du site. (Certaines pages reprennent des fichiers CSS déjà existant car cela ne créait pas de modifications trop importantes). Nous avons également utilisé quelques fonctions javascript. Enfin, nous nous sommes servis du langage PHP pour compléter les maquettes déjà codées et coder les pages et les fonctions avancées restantes. Nous avons ainsi pu obtenir des pages fonctionnelles, liées entre elles et avec des éléments cliquables (dirigeables). Nous avons travaillé avec la version 8.0 de PHP à l'aide de l'outil XAMPP ou MAMPP.

## II. Chargement des données

Afin de charger les données génomiques associées aux données des gènes et protéines , il faut tout d'abord créer le schéma de la base de données sous PostgreSQL en insérant le script “the\_last\_DB.sql”. Ensuite les données sont insérées dans la base de données à l'aide d'un script Python “Insert\_data.py” qui appelle les fonctions de parsing des fichiers fasta dans le script “parser\_files.py”. Les différents fichiers fasta sont regroupés dans un dossier data contenant les dossiers completegenome, pep\_nonannoté, cds\_aannoté, cdsfile, et peptidefile. Ensuite, il est nécessaire d'insérer le script SQL Insertion\_users.sql pour insérer des utilisateurs fictifs dans notre base de données et mettre à jour la table génome afin d'insérer les espèce et les souches. En effet, la base de données constitue l'élément clé du site dans la mesure où les données et le rôle de chaque utilisateur se doivent d'y être stockées afin d'accéder au contenu du site Web. Ainsi la majorité des pages php se doivent de contenir en en-tête les lignes de codes relatives à la connexion à la base de données (fichier “db\_info.ini” appelé à l'aide de “db\_utils.php”).

## Retour Expérience

Ce projet s'est avéré très intéressant en termes de contenu mais également en termes d'organisation. Il nous a en effet confronté à de nouveaux challenges auxquels nous n'avions pas encore fait face. Ainsi il s'agit d'un projet qui demande beaucoup de rigueur, d'organisation et de communication au sein du groupe afin de mener le projet à bien. Nous nous sommes organisés de manière à se voir au minimum une fois par semaine et discuter de l'avancement des uns et des autres, discuter des problèmes rencontrés, s'entraider pour la configuration et le débogage, définir les objectifs pour la fois suivante.

Cependant, au début du projet nous avons rencontré quelques soucis de cohésion. En effet nous travaillions toujours ensemble de manière hebdomadaire, mais pour autant nous n'avions pas défini clairement les rôles de chacun et la répartition des tâches ce qui a engendré des problèmes d'organisation au sein du groupe. Nous avons néanmoins vite surpassé cette phase avec beaucoup de communication entre les membres du groupe afin de mettre au clair les points qui posaient problèmes et de mettre en œuvre une nouvelle répartition plus adaptée. Le retour obtenu après le rendu 1 nous a également aidé à nous rebooster afin de repartir sur des bases solides et reprendre le travail avec davantage de sérieux et de motivation.

Nous avons choisi de travailler en binôme en se séparant selon les systèmes d'exploitation des machines que nous possédions. Sanâ et Lorraine ont ainsi travaillé ensemble sous Linux et Ilhem et Arnaud sous Windows. Il était ainsi plus facile de s'aider tant au niveau du code qu'au niveau des installations. On a ainsi réparti les tâches de sorte à décomposer le codage des différentes pages selon les binômes. Nous faisions des sessions de programmation par groupe (en binôme, trinôme) mais également seul afin d'avancer plus rapidement sur certaines parties du code. Le répertoire github était très régulièrement mis à jour et nous permettait ainsi d'avoir accès en temps quasi réel au travail réalisé par les autres membres du groupe et donc de suivre l'avancement.

## Sources

1. Sticky Navigation Bar in HTML CSS and JavaScript. *CodingLab | HTML CSS JavaScript Blogs - Free Source Code*. Available at:  
<https://www.codinglabweb.com/2021/08/sticky-navigation-bar-html-css.html>. Accessed December 10, 2021.
2. PHP: pg\_query - Manual. Available at:  
<https://www.php.net/manual/en/function.pg-query>. Accessed December 10, 2021.
3. Advanced Pagination < PHP | The Art of Web. Available at:  
<https://www.the-art-of-web.com/php/pagination/>. Accessed December 10, 2021.
4. PHP: file\_get\_contents - Manual. Available at:  
<https://www.php.net/manual/fr/function.file-get-contents.php>. Accessed December 10, 2021.
5. NCBI BLAST AMI (DEPRECATED) : Common URL API. Available at:  
<https://ncbi.github.io/blast-cloud/dev/api.html> and  
<https://github.com/AshokHub/BLASTphp>. Accessed December 10, 2021.
6. PHP: fputcsv - Manual. Available at:  
<https://www.php.net/manual/fr/function.fputcsv.php>. Accessed December 10, 2021.
7. How to Force download a File in PHP. *PHP Tutorial*. Available at:  
<https://www.phptutorial.net/php-tutorial/php-download-file/>. Accessed December 10, 2021.
8. Export Data from MySQL to CSV Using PHP. *iTech Empires* 2016. Available at:  
<https://www.itechempires.com/2016/05/export-data-from-mysql-to-csv-using-php/>. Accessed December 10, 2021.