

3/11/2021

Nous avons prévu un accès banques de données : bases de données du NCBI (National Center for Biotechnology Information) :

Assembly database, Gene Expression Omnibus, Protein database, EnsemblBacteria, Uniprot, ainsi qu'à BLAST pour l'alignement des séquences. Sur la page de chaque gène on pourra ainsi retrouver un accès à chacun de ces outils. L'utilisateur sera renvoyé sur la page web en question et pourra ainsi compléter les informations présentes sur la page d'annotation, à savoir approfondir l'analyse des protéines et gènes en question en particulier, lancer un alignement de séquence (via le BLAST), chercher des motifs particuliers ou créer une phylogénie par exemple.

Concrètement pour NCBI Blast qui est une ressource partagée pour trouver des similarités entre séquences. On retrouvera donc l'accès sur la page de chaque séquence à la fois au serveur BLAST, pour ce faire on pourra encoder la séquence FASTA directement ou le numéro d'accès (exemple : ASM584v2 ). Le passage se fera via une API BLAST à travers une interface https. L'API servira de portail entre l'application web et le serveur BLAST. Lors du transfert vers BLAST, on pourra également renseigner l'organisme duquel est issu le gène dans le champs de la requête prévu.

Pour Uniprot le principe est similaire, on pourra baser la recherche sur la séquence FASTA directement ou

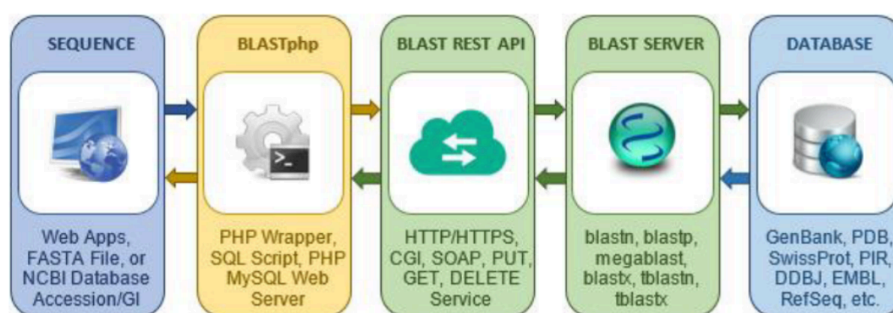


Figure 1. Remote execution of NCBI BLAST server through BLASTphp.

Lien de ces banques des données :

<https://www.ncbi.nlm.nih.gov/assembly> <https://www.ncbi.nlm.nih.gov/geo>

<https://www.ncbi.nlm.nih.gov/protein> <https://www.uniprot.org> <https://bacteria.ensembl.org>