

Université Mohamed KHEIDER Biskra
Faculté des Sciences Exactes et Sciences de la Nature et de le Vie
Département des Sciences de la Nature et de la Vie

1^{ère} Année mastère : Biochimie Fondamentale et Appliquée, Microbiologie et Biologie végétale.

Module : Bioinformatique

TP04 : Alignements de séquences Biologiques (BLAST)

Introduction

L'alignement est le processus par lequel 2 séquences sont comparées afin d'obtenir le plus de correspondances possibles entre les lettres qui les composent (nucléotides ou acides aminés).

- alignement global : comparaison de deux séquences sur toute leur longueur. Il permet de déterminer le degré de similarité de 2 séquences.
- alignement multiple : alignement global appliqué à plusieurs séquences à la fois. Il permet de rechercher des motifs conservés, prédire une structure, effectuer une étude phylogénétique.
- alignement local : comparaison de deux séquences seulement sur une portion de leur longueur (blocs). Il permet de comparer une séquence inconnue avec une banque de séquences connues.

Le logiciel Blast

Le mot Blast est un acronyme signifie (Basic local alignment search tool) qui est un algorithme de comparaison de séquences utilisé pour effectuer des recherches dans les banques en utilisant des alignements locaux. Il permet de faire des recherche de similarité à partir d'une séquence (protéique ou nucléique) sur toutes les séquences (ou un sous-ensemble de séquences) existantes. L'idée de base exploitée par l'algorithme est que les bons alignements (comparaisons) doivent contenir quelque part des petites régions très riches en identité.

L'utilisation du logiciel Blast

Vous pouvez utiliser ce logiciel directement en consultant le portail NCBI : <https://blast.ncbi.nlm.nih.gov/Blast.cgi>.

1. Soumettre une séquence d'ADN

- Vous allez utiliser BLAST pour comparer la séquence nucléique (''Gliadin, Accession number : KC660359'' en format FASTA) avec les séquences nucléotidiques de la base de données.
- Cliquez sur 'nucleotide blast' dans la section 'Basic BLAST'.
- Collez la séquence de Gliadin récupérée de la banque nucléotide
- Lancer une recherche BLAST pour la séquence ''Gliadin'', avec comme base de donnée nr/nt (vous pourrez choisir UniprotKB lorsqu'il s'agit d'un Blast protéique) et en vérifiant que la case 'Low complexity regions' est cochée. Attention, il faut lancer blastn et non megablast.
- Attendre le résultat de la recherche, ça peut prendre quelques dizaines de secondes. Vous accéderez automatiquement à la page des résultats.

2. Comparer une séquence d'ADN avec celles des bases de données

Vous voyez maintenant le résultat de la recherche BLASTn. La page résultat est divisée en 3 parties comme nous allons vu en cours (fichier attaché en format PDF) :

- Une vue graphique générale des séquences résultats avec différentes couleurs ;
- ensuite la liste des séquences avec leur score et leur E-value ;
- enfin, une vue plus détaillée, fournissant pour chaque séquence résultat, l'alignement avec notre séquence requête.

Revenez à la partie graphique. Notre séquence (Query) est représentée par la ligne épaisse rouge, graduée de 1 à 852. Le score de chacun des alignements est indiqué par une des 5 couleurs différentes. Plus le score est grand, plus la qualité est bonne et plus le pourcentage d'identité est élevé. Plusieurs segments de similarité trouvés dans une même séquence sont reliés par une ligne discontinue.

3. Déterminer l'identité des segments d'ADN

Utilisez votre curseur de souris pour vous placer sur une barre d'alignement colorée, vous verrez apparaître le nom de la séquence ainsi que le score d'alignement et la valeur d' E-value dans la zone de texte située au-dessus du graphique.

Si vous cliquez sur une barre, vous serez amené à l'alignement correspondant dans la 3e partie de la page.

Regardez les différentes séquences résultats obtenues et trouvez dans la liste le gène Gliadin. Rendez-vous à son alignement avec notre séquence requête.

Le segment aligné de notre séquence requête commence à la position 1 et finit à la position 852.

Questions

- Combien de séquence de la banque ressemblent à la notre (voir le nombre de 'hits') ?
- Est-ce que les alignements obtenus semblent pertinents d'un point de vue biologique ? justifiez !
- Est-ce que les séquences trouvées font partie de la famille des Gliadines ?

4. Chercher dans une banque de données protéique

- Copier la séquence protéique de l'insuline (récupérée à partir de la banque de données UniprotKB), retourner sur la page d'accueil de BLAST et choisissez cette fois 'protein blast'.
- Lancez la recherche avec comme base de donnée UniprotKB et la case 'Low complexity regions' cochée.