

## TP de Bioinformatique n°2: Alignement de séquences

### Objectifs du TP :

Comprendre le résultat du programme **BLAST**;

Utiliser un programme d'alignement multiple.

#### *Quelques "trucs" utiles :*

- Ctrl A permet de tout sélectionner sur une page
- Ctrl C permet de copier un texte sélectionné
- Ctrl V permet de le coller.
- Ctrl F permet de chercher un mot sur une page

### 1. Activité 1 : BLAST

Contexte :

En 1766 près de Villefranche de Rouergue, une jeune fille a été dévorée par la mystérieuse « bête du Gévaudan ». Quelle est la vraie nature de cette « bête » ? Loup ? Chien fou ? Psychopathe tueur en série ? Loup-Garou ?...

Les restes de la jeune victime sont encore conservés comme reliques dans la crypte de l'église. On a pu retrouver du sang séché sous ses ongles. Le laboratoire de la police scientifique a été chargé d'analyser ce sang séché. L'analyse a permis d'identifier et d'amplifier par PCR la séquence suivante :

>Sequence\_mystere

```
CAGCTGGAGACCAAACTCTAGGATCTAGGTCACCCCCTTGGCAGCAGAGCTGATTTATGGTGCCCCTGACAGCC  
TAATATCACCAAATTACCCACACAATGGAGAGAGGCTGGGCTGCCTCCTGGCAGTCGCTAAAGGACGCTGAG  
CCTGGGGCAGCCCTGCCGCTTCTGATCCCAGACCTCCTTGGGTCTGTGAGAGAGGGGGCTGGTGCAGACAG  
AGGCACACCCAGGCATGCCATCATCAACTGGGGGTATAAGTAGGAAGCCTTGGGGACAGCTCTGTACCCTGCA  
CTCGGGAGCCAGTAGCACTCGGACGACTACTTCTGTCCCACAAGCTGCAACAGAGCCAGCCATGAATTGCAG  
ACAGTTCTTGTCTCCACTGCAGCCGTGACAGCTCTGGGGGTGGTGGGGGCAGCAAGATGTTCTCCTGCAACC  
GCAGCAGCTTTTACGGGGCTAGTGGAGGAGGGGGCCGATTCAGCTCTTCTAGGAGCTTCGGAGGGGGGCAGCT  
CTGGGGCCTGTGGGAGGGGAGGTGGTGGCAGTTTTGGCTCTAGCTACGGTGGAGGATCTGGGGGCGGTTTTA  
GTGCTGGTAGCTTTGGGGGACATTCTAGGGGCTTCAGTGGTGGTTCTGGAGGAGGCTTTGGCGGTGGATTGG  
AGGAGGCTTTGGGGGCTTTAGTGGTGGTTCTGGAGGAGGCTTT
```

#### 1.1. Soumettre une séquence d'ADN

- Ouvrez le site <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
- Cliquez sur 'nucleotide blast' dans la section 'Basic BLAST'. Vous pouvez explorer les différentes possibilités offertes par BLAST en cliquant sur les icônes points d'interrogation, par exemple à côté du

menu déroulant 'Database', ou à côté de 'Enter an Entrez query to limit search'. Pour refermez la cadre d'information qui apparaît, il suffit de re-cliquer sur le point d'interrogation. Ouvrez les listes des paramètres en cliquant sur 'Algorithm parameters', et explorer toujours à l'aide des points d'interrogation les différentes options.

- Copier-coller cette séquence dans une fenêtre de requête de BLAST. Cochez la case Exclude Uncultured/environmental sample sequences. Lancer une requête de recherche de séquence parmi les banques de données (nucleotide collection nr nt) pour les séquences très similaires.
- Attendre le résultat de la recherche, ça peut prendre quelques dizaines de secondes. Vous accéderez automatiquement à la page des résultats

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

- Vous voyez maintenant le résultat de la recherche BLAST. La page résultat est divisée en 3 parties :
  1. Une vue graphique générale des séquences résultats avec différentes couleur ;
  2. ensuite la liste des séquences avec leur score et les % d'identité.
  3. enfin, une vue plus détaillée, fournissant pour chaque séquence résultat, l'alignement avec notre séquence requête.
- Revenez à la partie graphique. Notre séquence est représentée par la ligne épaisse rouge, graduée de .. à ..... notre séquence fait exactement .....nucléotides de long).
- 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.

### 1.3. Déterminer l'identité des segments d'ADN

- Combien de séquences de la banque ressemblent à la nôtre (voir le nombre de 'hits') ?
- 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 dans la zone de texte située au-dessus du graphique.
- Les trois meilleurs alignements fournis par BLAST correspondent à un ADN synthétique ainsi que ceux issus de deux espèces différentes. Indiquer ces espèces.
- D'après le résultat d'alignement (Rendez-vous aux alignements des 2 espèces avec notre séquence requête).
  - Combien de mutations existent-ils en comparant notre séquence avec la séquence NG\_008300.2
  - Combien de Gaps existent-ils en comparent notre séquence avec la séquence NG\_008300.2
  - Identifier l'espèce à laquelle appartient l'échantillon (nom latin et nom commun). Justifier votre réponse.

#### 1.4. Obtenir plus d'information sur une séquence

Cliquez sur le lien GenBank correspondant à la séquence codante donnant le meilleur score d'alignement. A partir de cette fiche et des liens qui lui sont associés, déduire : le nom du gène auquel appartient cette séquence ; la protéine codée ;

- Donnez les 6 premiers acides aminés de cette protéine.

#### 1.5. Télécharger le résultat d'un BLAST

- Télécharger les séquences FASTA des [AF000949.1](#), [NG\\_008300.2](#), [AC185986.3](#),

#### Activité 2 : Alignement avec MEGA 5

- 1) Ouvrir le logiciel MEGA 5.
- 2) Lancez un nouvel alignement : a-Align, b-Edit /Build alignement, c-create a new alignement, d-ok, e-DNA.
- 3) Copier-coller (CTRL+C-CTRL+V) les séquences téléchargées dans la fenêtre qui apparait.
- 4) Copier-coller la séquence Mystère avec les autre séquences.
- 5) Lancez un alignement, pour cela suivez les étapes : a-alignement, b- align by ClustalW
- 6) Téléchargez l'alignement en format FASTA en suivant les étapes a-DATA b- Export Alignement c-FASTA format.