

IFT3295 - Démo 1

11 Septembre 2019

Alignement de séquences

Lecture supplémentaire : <http://bit.ly/2eLMSzW>

Score d'un alignement de paire de séquences : pénalité de gaps

La pondération des gaps influence grandement l'alignement entre deux séquences. Certaines pondérations ont parfois plus de sens biologique que d'autres. Soit l'alignement A suivant :

```
S: ACGTTTACG-TAC-T  
T: ACC---ACGCTAGCT
```

En considérant les scores suivants :

- "match" : +1
- "mismatch" : -1
- "gap" : -1

Donnez les scores de l'alignement A pour chacune des pondérations de gaps suivantes :

1. pondération constante
2. pondération linéaire
3. pondération affine (utiliser une extension de gap $\sigma = -0.5$)
4. pondération convexe

Programmation dynamique et alignement pairwise de séquences

1. À quoi sert un alignement de séquences dans le contexte d'évolution de génomes ?
2. Alignement avec distance d'édition
 - (a) Donner les équations de récurrences ainsi que les conditions initiales pour construire la table de programmation dynamique pour l'alignement entre deux séquences S et T avec la distance d'édition.
 - (b) Construire la table de programmation dynamique pour les deux séquences suivantes : $S = CTGCGACA$ et $T = TGCAAG$
 - (c) Comment peut-on retrouver un alignement optimal de séquence à partir de la table ?
 - (d) Donner la complexité de l'algorithme d'alignement de paires de séquences avec la distance d'édition.
3. Que faut-il changer si on désire utiliser la similarité plutôt qu'une mesure de distance pour aligner deux séquences ?
4. Alignement local vs. global
 - (a) Quelle est la différence entre un alignement local et un alignement global ?
 - (b) Dans le cas d'un alignement local, la distance d'édition est-elle utile ? Justifiez
 - (c) Donnez les conditions initiales et les équations de récurrence de l'alignement local Smith-Waterman et de l'alignement global Needleman-Wunsch
5. *Recherche de motifs
6. *Problème de la plus longue sous-séquence commune entre deux séquences

Questions Générales

1. Quelle est la différence entre séquence protéique, ADN, ARN ?
2. À quoi sert le code génétique ?
3. Que désignent les termes suivants : Séquençage , Assemblage, Annotation, Homologie ?

Base de données biologiques

1. Format fastq et fasta
2. Blast sur NCBI : <https://blast.ncbi.nlm.nih.gov/Blast.cgi> de la protéine :

```
MALWMRLPLLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRRE  
AEDLQVGQVELGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN
```

3. Téléchargement de séquences homologues et alignement avec Clustal-Ω
<https://www.ebi.ac.uk/Tools/msa/clustalo/>.