

Alignements par paires de séquences protéiques avec l'algorithme de Smith-waterman

ELHAMRAOUI Zahra ELBARCHE F.Zahra

¹Université Mohammed V Rabat
Faculté de médecine et de pharmacie

mai , 2019

Plan

- 1 Introduction
- 2 alignement Local (Algorithme De Smith-Waterman)
- 3 Algorithmes et Packages
- 4 Démonstration de l'application
- 5 Conclusion

- 1 Introduction
- 2 alignement Local (Algorithme De Smith-Waterman)
- 3 Algorithmes et Packages
- 4 Démonstration de l'application
- 5 Conclusion

- En bioinformatique, l'opération d'alignement vise à identifier des zones communes à un groupe de k séquences.
- Des zones qui se ressemblent sont dites similaires ou homologues si elles dérivent d'un ancêtre commun.
- Un alignement sert à quantifier le degré de similarité ou de dissimilarité entre des séquences.

Trois Situations sont possibles pour une position donnée de l'alignement:

- Les caractères sont les mêmes: Identité
- Les caractères ne sont pas les mêmes: Substitution
- L'une des position est un espace: Insertion/ délétion

```
AAB24882      TYHMCQFHCRYVNNHSGEKLIECNERSKAFSCPSHLQCHKRRQIGETIHEHNQCGKAFPT 60
AAB24881      -----YECNQCCKAFAQHSSLKCHYRTHIGKEPYECNQCCKAFSK 40
               ****: .***: * *:*** *:***:..*****..

AAB24882      PSHLQYHRTHTGKPYECHQCGQAFKKCSLLQRHKRTHTGKPYE-CNQCCKAFAQ- 116
AAB24881      HSHLQCHKRTHTGKPYECNQCCKAFSQHGLLQRHKRTHTGKPYMNVIMVPLHNS 98
               **** *:*****:***:*.: .*****:
               *:*****:***:*.: .*****:
               *:*****:***:*.: .*****:
```

http://en.wikipedia.org/wiki/Sequence_alignment

identité

Insertion / Délétion

substitution

Quelles sont les régions de forte similarité entre les 2 séquences ?

Plan

- 1 Introduction
- 2 alignement Local (Algorithme De Smith-Waterman)**
- 3 Algorithmes et Packages
- 4 Démonstration de l'application
- 5 Conclusion

Définition

- L'alignement locale est un alignement de sous-séquences de A avec des sous-séquences de B.

Aligne seulement les régions dont le score est supérieur à un seuil donné .

Il est utilisé lorsque l'on veut aligner deux séquences de taille très différente .

Algorithme de Smith-waterman

- L'algorithme de Smith-Waterman est un algorithme optimal qui donne un alignement correspondant au meilleur score possible de correspondance entre les acides aminés ou les nucléotides des deux séquences. Le calcul de ce score repose sur l'utilisation de matrices de similarité ou matrices de substitution.

Plan

- 1 Introduction
- 2 alignement Local (Algorithme De Smith-Waterman)
- 3 Algorithmes et Packages**
- 4 Démonstration de l'application
- 5 Conclusion

"Algorithmes et fonctions " utilisées

- **pairwiseAlignment()** : Résout l'alignement global (Needleman-Wunsch), l'alignement local (Smith-Waterman) et les problèmes d'alignement (sans extrémités).
- **split_word()** : permet de diviser un mot en des lettres séparés

- **DotPlot()**: Permet la représentation graphique simple des résidus identiques entre les deux séquences.
- **heatmap()** : Permet la représentation graphique de données dans laquelle les valeurs individuelles contenues dans une matrice sont représentées en couleurs.

"Packages" utilisées

- **Biostrings** : Conteneurs d'algorithmes permettant une manipulation rapide de grandes séquences biologiques ou d'ensembles de séquences. contient la fonction `PairwiseAlignment()`

- **seqinr** : Permet une manipulation efficace des chaînes biologiques contient la fonction `splitseq()` qui fait divisé une séquence en sous séquences
- **stringr**: se concentre sur les fonctions de manipulation de chaînes les plus importantes et les plus utilisées contient la fonction `str_split(x, pattern)`

- **shiny** : permet de créer un panneau dont le contenu est absolument positionné , afin de réaliser des interfaces graphiques .

Plan

- 1 Introduction
- 2 alignement Local (Algorithme De Smith-Waterman)
- 3 Algorithmes et Packages
- 4 Démonstration de l'application**
- 5 Conclusion

Plan

- 1 Introduction
- 2 alignement Local (Algorithme De Smith-Waterman)
- 3 Algorithmes et Packages
- 4 Démonstration de l'application
- 5 Conclusion**

- L'alignement des séquences biologiques est probablement le plus important dans le domaine de la bioinformatique. Leur but ultime est de déterminer la similarité entre différentes séquences. L'alignement de la séquence a donc un grand nombre d'applications :
 - Recherche de gènes
 - Assemblage de la séquence du génome
 - identification de gènes homologues