

Démos déplacées sur les créneaux du cours :

- 12 novembre (démon mardi, cours 3h mercredi)
+ *pas de office hours le 14 novembre*
- 19 novembre (démon mardi, rien le mercredi)
- 26 novembre (démon mardi, cours 3h mercredi)
- Reste des cours/démos inchangé

TP3 à rendre avant le jeudi 21 novembre (mercredi soir 23h59 dernier délai)

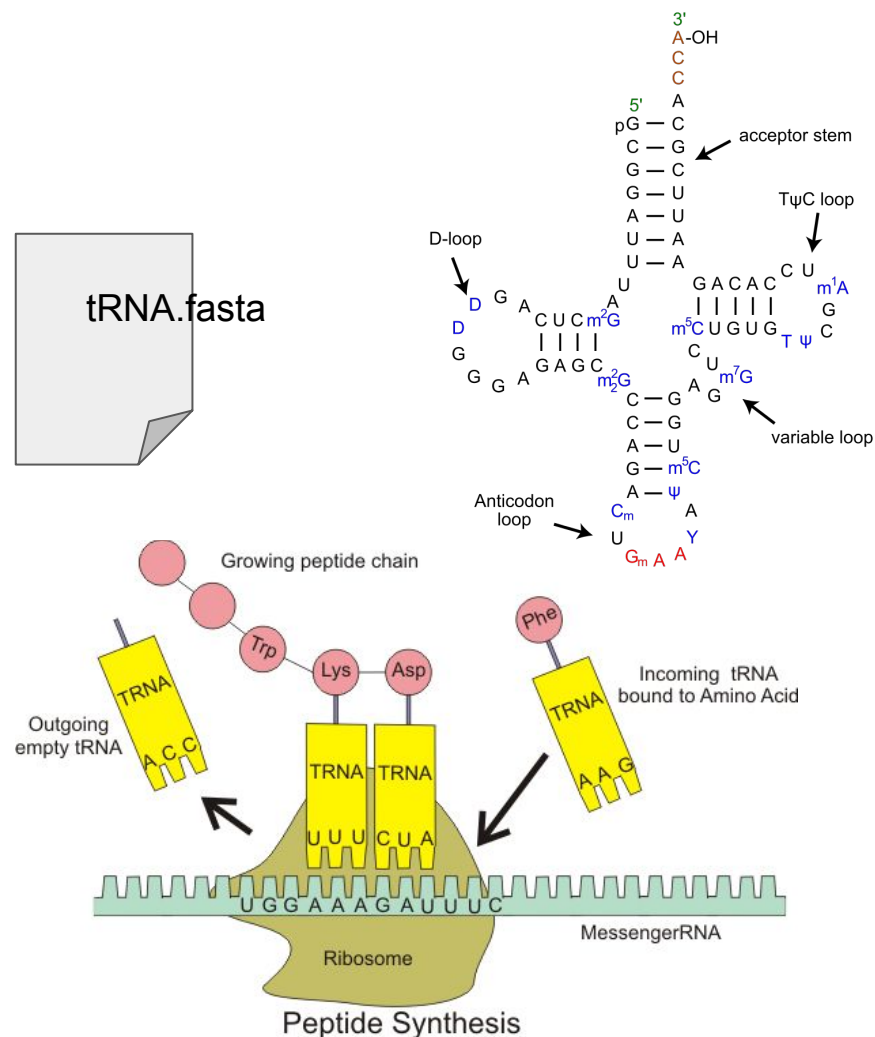
Présentation du TP3

Etapes d'implémentation

1. Banques de données
2. Graines et k-mers
3. Recherche exactes des mots de taille k
4. Heuristique pour extension des HSPs
5. Fusion des HSPs chevauchant
6. Statistiques sur les HSPs et cutoff
7. Alignement SW et output

Mise en pratique

1. Test du programme sur plusieurs séquences query
2. Dédire la nature de ces séquences inconnues
3. Comparer avec le vrai BLAST
4. Influence de la longueur des seeds
5. Autre type de graine



Présentation du TP3

Recherche de HSP = fragments similaires
entre la séquence requête et les séquences
de bases de données

Etapas d'implémentation

1. Banques de données
2. **Graines et k-mers**
3. Recherche exactes des mots de taille k
4. Heuristique pour extension des HSPs
5. Fusion des HSPs chevauchant
6. Statistiques sur les HSPs et cutoff
7. Alignement SW et output

Mise en pratique

1. Test du programme sur plusieurs séquences query
2. Dédire la nature de ces séquences inconnues
3. Comparer avec le vrai BLAST
4. Influence de la longueur des seeds
5. Autre type de graine

Séquence requête



...TACGTATGTCTCTAGTCGTGC...

|||||
1111111111
ATGTCTCTAGT

Graine de taille 11
avec des matches
à chaque position

...TACGTATCTTTTCATGTCGTGC...

|| | ||
11010110011
ATGTCTCTAGT

Autre type de
graine de taille 11

Présentation du TP3

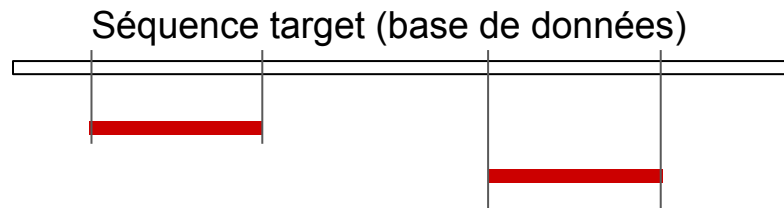
Etapes d'implémentation

1. Banques de données
2. Graines et k-mers
3. Recherche exactes des mots de taille k
4. Heuristique pour extension des HSPs
5. Fusion des HSPs chevauchant
6. Statistiques sur les HSPs et cutoff
7. Alignement SW et output

Mise en pratique

1. Test du programme sur plusieurs séquences query
2. Déduire la nature de ces séquences inconnues
3. Comparer avec le vrai BLAST
4. Influence de la longueur des seeds
5. Autre type de graine

Recherche des HSP initiaux :
Occurences des kmers de la séquence
requête dans les séquences de la bases de
données



Présentation du TP3

Etapes d'implémentation

1. Banques de données
2. Graines et k-mers
3. Recherche exactes des mots de taille k
4. **Heuristique pour extension des HSPs**
5. Fusion des HSPs chevauchant
6. Statistiques sur les HSPs et cutoff
7. Alignement SW et output

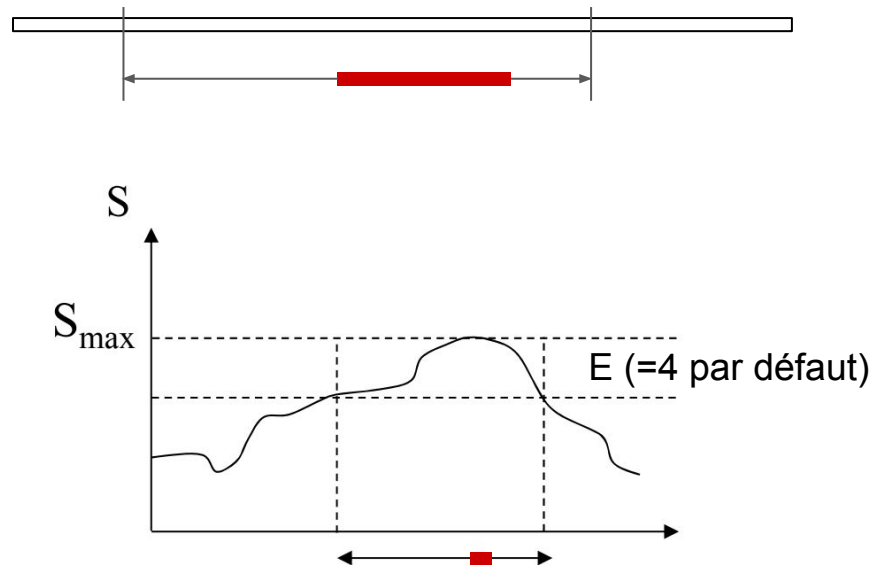
Mise en pratique

1. Test du programme sur plusieurs séquences query
2. Dédire la nature de ces séquences inconnues
3. Comparer avec le vrai BLAST
4. Influence de la longueur des seeds
5. Autre type de graine

Extension des HSPs :

Essayer d'étendre l'alignement des deux cotés jusqu'à ce que le score cumulé de l'alignement chute trop.

(score: +5 match, -4 mismatch)



Présentation du TP3

Etapes d'implémentation

1. Banques de données
2. Graines et k-mers
3. Recherche exactes des mots de taille k
4. Heuristique pour extension des HSPs
5. **Fusion des HSPs chevauchant**
6. Statistiques sur les HSPs et cutoff
7. Alignement SW et output

Mise en pratique

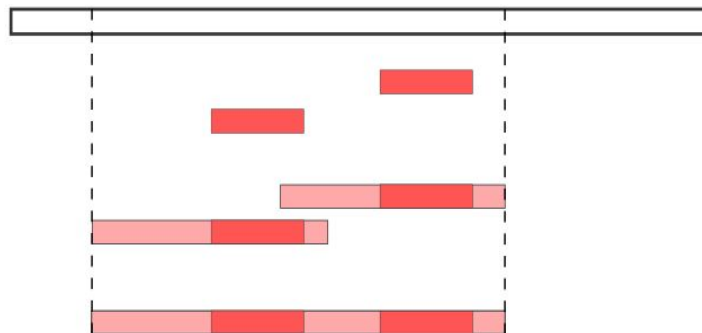
1. Test du programme sur plusieurs séquences query
2. Déduire la nature de ces séquences inconnues
3. Comparer avec le vrai BLAST
4. Influence de la longueur des seeds
5. Autre type de graine

Séquence Target

Match exact
de kmer

Extension
de match

Fusion de
HSP chevauchant



Présentation du TP3

Etapes d'implémentation

1. Banques de données
2. Graines et k-mers
3. Recherche exactes des mots de taille k
4. Heuristique pour extension des HSPs
5. Fusion des HSPs chevauchant
- 6. Statistiques sur les HSPs et cutoff**
7. Alignement SW et output

Mise en pratique

1. Test du programme sur plusieurs séquences query
2. Déduire la nature de ces séquences inconnues
3. Comparer avec le vrai BLAST
4. Influence de la longueur des seeds
5. Autre type de graine

bitscore:

$$B = \text{round}\left(\frac{\lambda S - \ln(K)}{\ln(2)}\right)$$

e-value:

$$e = mn2^{-B}$$

⇒ Seuil de signification pour récupérer les HSPs pertinents

Présentation du TP3

Etapes d'implémentation

1. Banques de données
2. Graines et k-mers
3. Recherche exactes des mots de taille k
4. Heuristique pour extension des HSPs
5. Fusion des HSPs chevauchant
6. Statistiques sur les HSPs et cutoff
7. **Alignement SW et output**

Mise en pratique

1. Test du programme sur plusieurs séquences query
2. Dédurre la nature de ces séquences inconnues
3. Comparer avec le vrai BLAST
4. Influence de la longueur des seeds
5. Autre type de graine

```
>I|gat|Mesostigma_viride score: 20.0 ident: 81.25
AGTTGGTTAGA-GTATACGCCTGATAAGCGTA
AGTCGGCTA-ACGCATACGCTTGATAAGCGTA
```

```
# Best HSP score:78.00, bitscore:24.00, evalue: 7.67e-02
13  CGCATACGCTTGATAAGCGTA   33
23  AGTATACGCCTGATAAGCGTA   43
```

```
-----
Total : 1
```

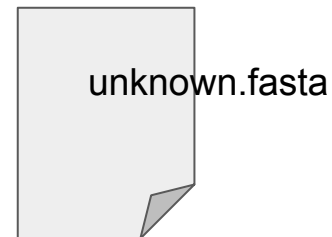

Présentation du TP3

Etapes d'implémentation

1. Banques de données
2. Graines et k-mers
3. Recherche exactes des mots de taille k
4. Heuristique pour extension des HSPs
5. Fusion des HSPs chevauchant
6. Statistiques sur les HSPs et cutoff
7. Alignement SW et output

Mise en pratique

1. **Test du programme sur plusieurs séquences query**
2. **Déduire la nature de ces séquences inconnues**
3. **Comparer avec le vrai BLAST**
4. **Influence de la longueur des seeds**
5. **Autre type de graine**



Pour chaque séquence inconnue :

- Sortie de votre PLAST (alignements et HSPs trouvés et triés)
- Nature des séquences ?
- Sortie de BLAST et comparaison