

Bioinformatique et biologie computationnelle

Groupe b

Détecter la bactérie E.coli dans l'eau

Eléonore D'agostino & Yannick Widmer

24.04.2016

Aperçu de la présentation

- Rappel et Résumé
- Réalisation
- Résultats
- Conclusion

Rappel

- Le but de notre projet est de concevoir un test pour détecter la bactérie E.coli dans l'eau
- On veut un test facile à utiliser, avec les meilleurs taux de détection possibles

Résumé

- S'il y a des traces de E.Coli dans l'eau elle n'est pas potable
- On cherche des épitopes de la protéine OmpF Porin de E. Coli, que l'on mettra dans un test
- On utilise un test appelé ELISA, qui contient une bande qui change de couleur si les anticorps dedans détectent leur épitope correspondant

Réalisation étape 1

- On obtient toutes les séquences d'E.Coli contenant la protéine OmpF Porin sur GenBank (1279 résultats)

```
query = "ompf"[gene] AND "Escherichia coli"[porgn]
```



NCBI
GenBank

Réalisation étape 2

- On garde que les parties correspondant à OmpF Porin et on les écrits dans un fichier.csv

```
ompf = 0
translation = ""
prot = ""
for f in features:
    if f['GBQualifier_name'] == 'gene' and f['GBQualifier_value'].lower() == "ompf":
        ompf = 1
    if f['GBQualifier_name'] == 'translation':
        translation = f['GBQualifier_value']
    if f['GBQualifier_name'] == "protein_id" :
        prot = f['GBQualifier_value']
    if translation and prot and ompf:
        csv.write("\\" + prot + "\",\\" + translation + "\\" + "\n")
        break
```

Réalisation étape 3

- On vérifie les résultats de notre CSV. Script => 2 séquences partielles.

```
C:\Users\Youn\Mes documents\bbc\bbcislife-master>python Liste_wrong_protein.py  
sep=  
Protein_ID  
EHX66884.1  
AAP13245.1
```


Réalisation étape 4

- On crée un fichier .fasta à partir du fichier csv. Puis on alignes les séquences



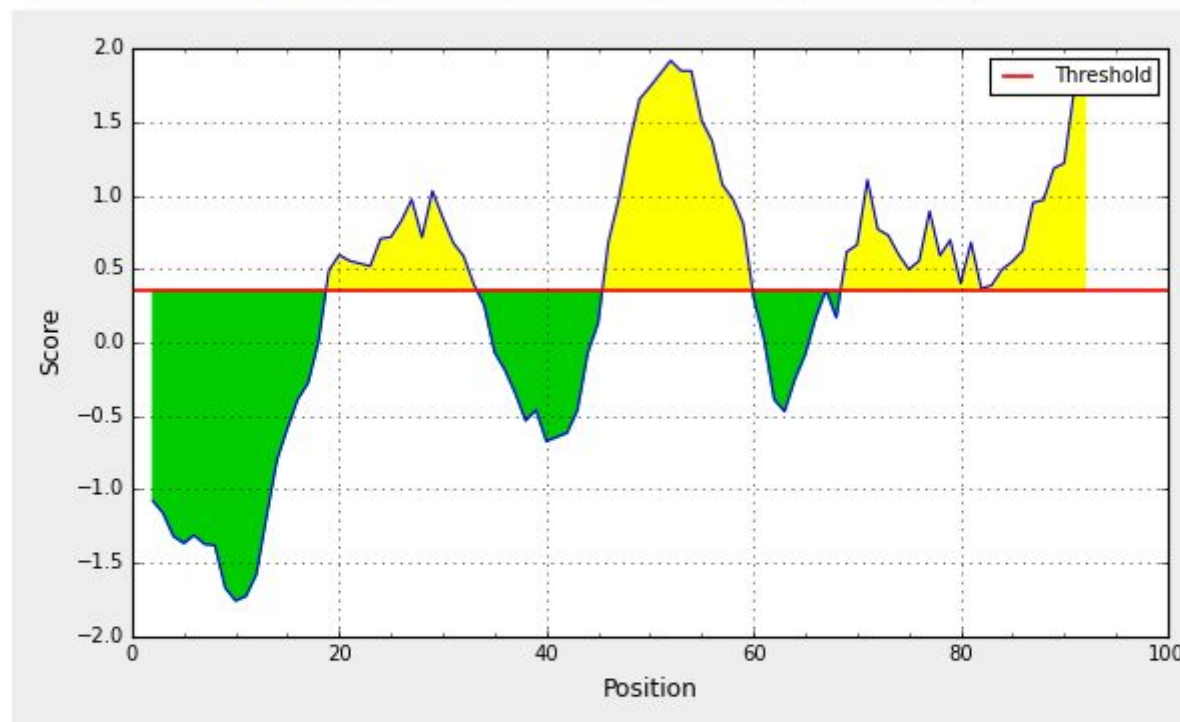
Réalisation étape 5

- On teste chaque région pour la présence d'épitopes potentiels

Input Sequences

1 MKRNILAVIV PALLIAGTAN AAEIYNKDGN KVDLYGKAVG LHYFSKNGE NSYGGNGDMT
61 YARLGFKGET QINSDLTGYG QWEYNFQGN SE

Center position: 4 Window size: 7 Threshold: 0.35 Recalculate



Average: 0.277 Minimum: -1.761 Maximum: 1.921

Predicted peptides:

No.	Start	End	Peptide	Length
1	19	33	ANAAEIYNKDGNKVD	15
2	46	59	KNGENSYGGNGDM	14
3	67	67	K	1

Réalisation étape 6

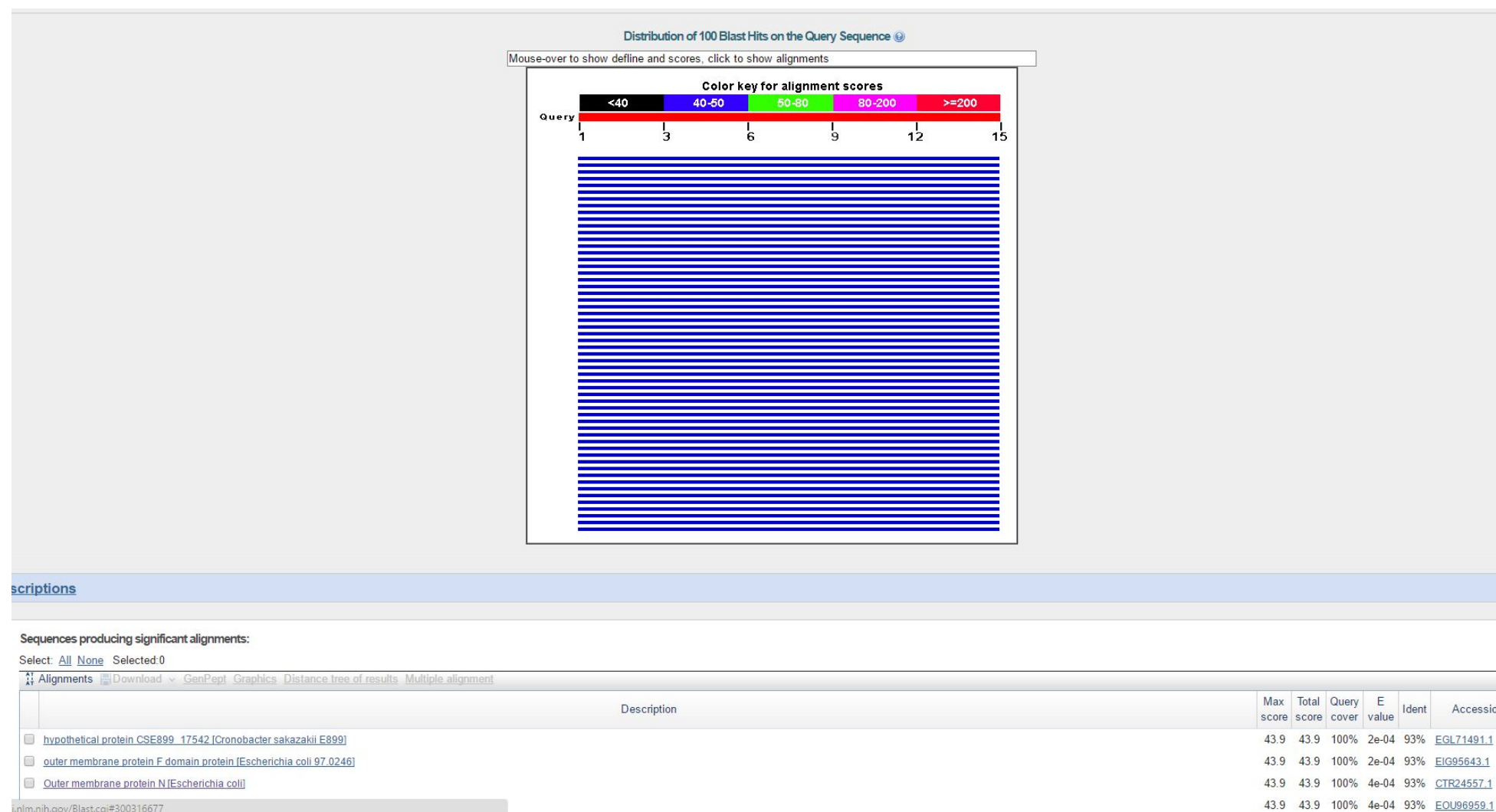
- On vérifie le nombre de cas présent dans la séquence :

```
C:\Users\Youn\Mes documents\bbc\bbcislife-master>python compare.py  
15  
1226
```

- Cette séquence est donc présente à 96% dans toutes les souches de E. Coli.

Réalisation étape 7

- Il manque la vérification d'autres espèces avec les séquences sélectionnées.



Résultats

- Sur 1279 séquences, on a dû en jeter deux (séquences partielles)
- On a trouvé 8 séquences intéressantes à tester, qui ont amené à 5 épitopes potentiels

Epitope	Longueur	Sensibilité
ANAAEIYNKDG NKVD	15	96.00%
KGNGENSYGGNGDM	14	86.69%
GADAQTGNK	9	93.58%
DTARRSNGDGVGGSISY	17	95.54%
GKKAEQWA	8	98.82%

Conclusion

- Longue séquence \neq bons épitopes
- Compromis pour trouver les épitopes
- Difficile de calculer les faux positifs

CONCLUSION

Vous trouvez les scripts ici : <http://younTheory/bbcislife>

Question

Questions

