

Rapport TP3 IFT3295

Asmaa Zohra Skou, 20217232

05 Décembre, 2024

1 Implémentation PLAST

Pour rouler le code, il suffit d'enregistrer les fichiers unknown.fasta et tRNAs.fasta dans le même directory que le fichier Plast.py. La commande que vous devez rentrer dans le terminal est sous la forme :

```
python3 plast.py -i AGCGGGGTAGAGGAATTGGTTTACTCATCAGGCTCATGACCTGAA-  
GACTGCAGGTTCTGAATCCTGT CCCCGCCT -db tRNAs.fasta -E 4 -ss 0.001 -seed 1111111111
```

* utiliser python3 dans la commande (MacOS)

* utiliser python dans la commande (Windows)

2 Mise en pratique

2.1 Unknown.fasta

Seq 1 : *Malus domestica*

```
>M|cat|Carica_papaya
# Best HSP score:255.00, bitscore:73.00, evalue: 2.45e-16
22 ACTCATCAGGCTCATGACCTGAAGACTGCAGGTTCTGAATCCTGTCCCCGCCT 73
22 ACTCATCAGGCTCATGACCTGAAGACTGCAGGTTCTGAATCCTGTCCCCGCCT 73
-----
>M|cat|Oryza_sativa_Japonica_Group
# Best HSP score:255.00, bitscore:73.00, evalue: 2.45e-16
22 ACTCATCAGGCTCATGACCTGAAGACTGCAGGTTCTGAATCCTGTCCCCGCCT 73
22 ACTCATCAGGCTCATGACCTGAAGACTGCAGGTTCTGAATCCTGTCCCCGCCT 73
-----
>M|cat|Vitis_vinifera_2
# Best HSP score:195.00, bitscore:57.00, evalue: 1.61e-11
34 CATGACCTGAAGACTGCAGGTTCTGAATCCTGTCCCCGCCT 73
34 CATGACCTGAAGACTGCAGGTTCTGAATCCTGTCCCCGCCT 73
-----
>M|cat|Arabidopsis_thaliana
# Best HSP score:120.00, bitscore:36.00, evalue: 3.37e-05
22 ACTCATCAGGCTCATGACCTGAAGA 46
22 ACTCATCAGGCTCATGACCTGAAGA 46
-----
>M|cat|Arabidopsis_thaliana
# Best HSP score:105.00, bitscore:32.00, evalue: 5.40e-04
50 CAGGTTCTGAATCCTGTCCCCGC 71
50 CAGGTTCTGAATCCTGTCCCCGC 71
-----
Total : 5
```

Figure 1: Output pour seq 1: *Malus domestica*

Les 2 premiers résultats sont identiquement les meilleurs avec un score HSP de 255 , un bitscore de 73 et un e-value de 2.45e-16. Les résultats les plus proches de *Malus domestica* sont donc **Carica papaya et Oryza sativa Japonica Group**.

Seq 2 : Nephroselmis olivacea

```
>R|tcg|Marchantia_polymorpha
# Best HSP score:135.00, bitscore:40.00, evalue: 2.11e-06
44 TGGTCACAGGTTCAAATCCTGTAGGATG 71
44 TGGTCACAGGTTCAAATCCTGTAGGATG 71
-----
>R|tct|Mesostigma_viride
# Best HSP score:110.00, bitscore:33.00, evalue: 2.70e-04
45 GGTCACAGGTTCAAATCCTGTAG 67
45 GGTCACAGGTTCAAATCCTGTAG 67
-----
>R|tct|Marchantia_polymorpha
# Best HSP score:110.00, bitscore:33.00, evalue: 2.70e-04
45 GGTCACAGGTTCAAATCCTGTAG 67
45 GGTCACAGGTTCAAATCCTGTAG 67
-----
Total : 3
```

Figure 2: Ouput pour seq 2 : Nephroselmis olivacea

Le premier résultat est le meilleur avec un score HSP de 135, un bitscore de 40 et un evalue de 2.11e-06. Le résultat le plus proche de Nephroselmis olivacea est donc **Marchantia polymorpha**.

Seq 3 : Phoenix dactylifera

```
>P|tgg|Oryza_sativa_Japonica_Group
# Best HSP score:135.00, bitscore:40.00, evalue: 2.11e-06
46 GTCACGGGTTCAAATCCTGTATCCCTA 73
46 GTCACGGGTTCAAATCCTGTATCCCTA 73
-----
>P|tgg|Sorghum_bicolor
# Best HSP score:135.00, bitscore:40.00, evalue: 2.11e-06
46 GTCACGGGTTCAAATCCTGTATCCCTA 73
46 GTCACGGGTTCAAATCCTGTATCCCTA 73
-----
Total : 2
```

Figure 3: Ouput pour seq3 : Phoenix dactylifera

Les 2 résultats ont exactement les mêmes valeurs qui sont un score HSP de 135, un bitscore de 40 et un evalue de 2.11e-06. Les résultats les plus proches de Phoenix dactylifera sont donc **Oryza sativa Japonica Group et Sorghum bicolor**.

Seq 4 : Chara vulgaris

```
>R|tct|Marchantia_polymorpha
# Best HSP score:255.00, bitscore:73.00, evalue: 2.45e-16
16 TGGATAGAGCAACAACCTTCTAAGTTGAAGGTCACAGGTTCAAATCCTGTAG 67
16 TGGATAGAGCAACAACCTTCTAAGTTGAAGGTCACAGGTTCAAATCCTGTAG 67
-----
>R|tcg|Marchantia_polymorpha
# Best HSP score:135.00, bitscore:40.00, evalue: 2.11e-06
45 GGTCACAGGTTCAAATCCTGTAGGATGC 72
45 GGTCACAGGTTCAAATCCTGTAGGATGC 72
-----
>R|tct|Mesostigma_viride
# Best HSP score:115.00, bitscore:34.00, evalue: 1.35e-04
44 AGGTCACAGGTTCAAATCCTGTAG 67
44 AGGTCACAGGTTCAAATCCTGTAG 67
-----
Total : 3
```

Figure 4: Seq 4 : Chara vulgaris

Le meilleur résultat pour Chara vulgaris est le **Marchantia polymorpha** avec un score HSP de 255, un bitscore de 73 et un evalue de 2.45e-16.

2.2 La nature de chacune des séquences

Séquence 1 : Malus domestica

Je déduis que l'acide aminé de la séquence 1 est "M,méthionine" et l'anticodon est "cat" considérant que les 5 résultats obtenus pour cette séquence ont la même nature.

Séquence 2 : Nephroselmis olivacea

Je déduis que l'acide aminé de la séquence 2 est "R,Arginine" et l'anticodon est "tcg" considérant que le meilleur résultat a la même nature. Cependant il se pourrait que l'anticodon soit "tct" considérant les 2 autres résultats.

Séquence 3 : Phoenix dactylifera

Je déduis que l'acide aminé de la séquence 3 est "P,Proline" et l'anticodon est "tgg" considérant les 2 résultats obtenus avec PLAST.

Séquence 4 : Chara vulgaris

Je déduis que l'acide aminé de la séquence 4 est "R,Arginine" et l'anticodon est "tct" considérant le meilleur résultat donné par PLAST, cependant il se pourrait que l'anticodon soit "tcg" si je prends en compte les 2 autres résultats.

2.3 Les résultats avec BLAST

Séquence 1 : "Malus domestica genome assembly, organelle: mitochondrion" Score : 137 et e-value = 1e-28

Séquence 2 : "Nephroselmis olivacea mitochondrion, complete genome" Score: 137 et e-value = 1e-28

Séquence 3 : "Phoenix dactylifera cultivar Khanezi isolate K2 mitochondrion, complete genome" Score: 137 et e-value = 1e-28

Séquence 4 : "Chara vulgaris mitochondrion, complete genome" Score: 137 et e-value = 1e-28

Si je compare les résultats obtenus avec PLAST et BLAST, je peux confirmer que l'outil BLAST est énormément plus précis que PLAST. Les résultats obtenus par PLAST sont des organismes proches de la séquence recherchée (ex: la séquence recherché : Nephroselmis olivacea = algue verte, le résultat de PLAST : Marchantia polymorpha = une mousse, hépatique). Contrairement à PLAST, l'outil BLAST donne comme résultat l'espèce recherchée exacte (un match de 100 pourcents avec un score HSP parfait et un evalue bas).

2.4 Impact des graines plus longues sur PLAST (vitesse, précision, sensibilité)

Je n'ai pas implémenté cette fonctionnalité cependant je déduis que si j'utilise des graines plus longues PLAST va être plus rapide considérant que ces dernières réduisent le nombre de correspondances initiales possibles dans les séquences (moins extensions). PLAST serait aussi plus précis vu que les longues graines réduisent le risque d'alignements non significatifs cependant PLAST serait moins sensible vu que les longues graines pourraient manquer des alignements réels qui contiennent des petites variations.

Les effets des graines plus courtes sont différents. PLAST serait plus lent vu que ces dernières génèrent beaucoup plus de correspondances initiales (plus d'extensions et d'alignements non significatifs). PLAST serait moins précis vu que l'utilisation d'une graine plus courte pourrait engendrer des correspondances dans des régions non significatives (trop d'alignements inutiles). Néanmoins, PLAST serait beaucoup plus sensible vu que les graines plus courtes peuvent mieux détecter des alignements avec des indels.