

Mise en pratique

Question 1 :

```
=====
Résultats pour : X|???|Malus_domestica
=====
```

```
-- Resultats PLAST --
```

```
Sequence : M|cat|Carica_papaya
Bitscore : 73
E-value : 2.4599721436572523e-16
Input : [21:73]
DB : [21:73]
Align input : TACTCATCAGGCTCATGACCTGAAGACTGCAGGTTGGAATCCTGTCCCCGCCT
Align db : GACTCATCAGGCTCATGACCTGAAGACTGCAGGTTGGAATCCTGTCCCCGCCT
```

```
Sequence : M|cat|Oryza_sativa_Japonica_Group
Bitscore : 73
E-value : 2.4599721436572523e-16
Input : [21:73]
DB : [21:73]
Align input : TACTCATCAGGCTCATGACCTGAAGACTGCAGGTTGGAATCCTGTCCCCGCCT
Align db : AACTCATCAGGCTCATGACCTGAAGACTGCAGGTTGGAATCCTGTCCCCGCCT
```

```
Sequence : M|cat|Vitis_vinifera_2
Bitscore : 71
E-value : 9.83988857462981e-16
Input : [21:73]
DB : [21:73]
Align input : TACTCATCAGGC
Align db : GACTCATCAGGC
```

```
Sequence : M|cat|Arabidopsis_thaliana
Bitscore : 66
E-value : 3.148764343881283e-14
Input : [21:71]
DB : [21:71]
Align input : TACTCATCAGGCTCATGACCTGAAGA
Align db : AACTCATCAGGCTCATGACCTGAAGA
```

```
Sequence : M|cat|Bracteacoccus_minor_2
Bitscore : 32
E-value : 0.0005409535951912403
Input : [11:40]
DB : [10:39]
Align input : GGAATTGGTTTACTCATCAGGCTCATGACC
Align db : AGTAGTGGTTAACTCATGGGCTCATGACC
```

```
=====
Résultats pour : X|???|Phoenix_dactylifera
=====
```

```
-- Resultats PLAST --
```

```
Sequence : P|tgg|Oryza_sativa_Japonica_Group
Bitscore : 40
E-value : 2.1130999812157825e-06
Input : [43:73]
DB : [43:73]
Align input : GAGGTCACGGGTTCAAATCCTGTATCCCTA
Align db : AATGTCACGGGTTCAAATCCTGTATCCCTA
```

```
Sequence : P|tgg|Sorghum_bicolor
Bitscore : 40
E-value : 2.1130999812157825e-06
Input : [43:73]
DB : [43:73]
Align input : GAGGTCACGGGTTCAAATCCTGTATCCCTA
Align db : AATGTCACGGGTTCAAATCCTGTATCCCTA
```

```
Sequence : P|tgg|Vitis_vinifera
Bitscore : 38
E-value : 8.45239992486313e-06
Input : [43:73]
DB : [43:73]
Align input : GAGGTCACGGGTTTC
Align db : AATGTCACGGGTTTC
```

```
=====
Résultats pour : X|???|Nephroselmis_olivacea
=====
```

```
-- Resultats PLAST --
```

```
Sequence : R|tcg|Marchantia_polymorpha
Bitscore : 85
E-value : 6.005791366350714e-20
Input : [0:71]
DB : [0:71]
Align input : ACATCCTTAGCTCAGT
Align db : GCATTCTTAGCTCAGT
```

```
Sequence : R|tct|Mesostigma_viride
Bitscore : 53
E-value : 2.579467750507547e-10
Input : [0:43]
DB : [0:43]
Align input : ACATCCTTAGCTCAGTAGGATAGAGCAACAGCCTTCTAAGCTGG
Align db : ACATTCTTAGCTCAGTTGGATAGAGCAACAGCCTTCTAAGCTGT
```

```
Sequence : R|tct|Marchantia_polymorpha
Bitscore : 37
E-value : 1.690479984972626e-05
Input : [0:29]
DB : [0:29]
Align input : ACATCCTTAGCTCAGT
Align db : GCATTCTTAGCTCAGT
```

```
=====
Résultats pour : X|???|Phoenix_dactylifera
=====
```

```
=====
Résultats pour : X|???|Chara_vulgaris
=====
```

```
-- Resultats PLAST --
```

```
Sequence : R|tct|Marchantia_polymorpha
Bitscore : 98
E-value : 7.331288289002336e-24
Input : [0:73]
DB : [0:73]
Align input : GCATTCTTAGCTCAGTTGGATAGAGCAACACCTTCTAAGTTGAAGGTCACAGGTTCAAACTCTGTAGGATGCT
Align db : GCATTCTTAGCTCAGTTGGATAGAGCAACACCTTCTAAGTTGAAGGTCACAGGTTCAAACTCTGTAGGATGCG
```

```
Sequence : R|tcg|Marchantia_polymorpha
Bitscore : 95
E-value : 5.865030631201869e-23
Input : [0:73]
DB : [0:73]
Align input : GCATTCTTAGCTCAGTTGGATAGAGCAACACCTTCTAAGTTGAAGGTCACAGGTTCAAACTCTGTAGGATGCT
Align db : GCATTCTTAGCTCAGTTGGATAGAGCAACACCTTCTAAGTTGAAGGTCACAGGTTCAAACTCTGTAGGATGCG
```

```
Sequence : R|tct|Mesostigma_viride
Bitscore : 48
E-value : 7.331288289002336e-24
Input : [0:73]
DB : [0:73]
Align input : GCATTCTTAGCTCAGTTGGATAGAGCAACACCTTCTAAGTTGAAGGTCACAGGTTCAAACTCTGTAGGATGCT
Align db : GCATTCTTAGCTCAGTTGGATAGAGCAACACCTTCTAAGTTGAAGGTCACAGGTTCAAACTCTGTAGGATGCG
```

```
Sequence : R|tcg|Marchantia_polymorpha
Bitscore : 95
E-value : 5.865030631201869e-23
Input : [0:73]
DB : [0:73]
Align input : GCATTCTTAGCTCAGTTGGATAGAGCAACACCTTCTAAGTTGAAGGTCACAGGTTCAAACTCTGTAGGATGCT
Align db : GCATTCTTAGCTCAGTTGGATAGAGCAACACCTTCTAAGTTGAAGGTCACAGGTTCAAACTCTGTAGGATGCG
```

```
E-value : 7.331288289002336e-24
Input : [0:73]
DB : [0:73]
Align input : GCATTCTTAGCTCAGTTGGATAGAGCAACACCTTCTAAGTTGAAGGTCACAGGTTCAAACTCTGTAGGATGCT
Align db : GCATTCTTAGCTCAGTTGGATAGAGCAACACCTTCTAAGTTGAAGGTCACAGGTTCAAACTCTGTAGGATGCG
```

```
Sequence : R|tcg|Marchantia_polymorpha
Bitscore : 95
E-value : 5.865030631201869e-23
Input : [0:73]
DB : [0:73]
Align input : GCATTCTTAGCTCAGCTGGATAGAGCAACAACCTTCTAAGTTGAAGGTCACAGGTTCAAATCCTGTAGGATGCT
E-value : 7.331288289002336e-24
Input : [0:73]
DB : [0:73]
Align input : GCATTCTTAGCTCAGCTGGATAGAGCAACAACCTTCTAAGTTGAAGGTCACAGGTTCAAATCCTGTAGGATGCT
Align db : GCATTCTTAGCTCAGTTGGATAGAGCAACAACCTTCTAAGTTGAAGGTCACAGGTTCAAATCCTGTAGAATGCG
```

```
Sequence : R|tcg|Marchantia_polymorpha
Bitscore : 95
E-value : 5.865030631201869e-23
E-value : 7.331288289002336e-24
Input : [0:73]
DB : [0:73]
Align input : GCATTCTTAGCTCAGCTGGATAGAGCAACAACCTTCTAAGTTGAAGGTCACAGGTTCAAATCCTGTAGGATGCT
Align db : GCATTCTTAGCTCAGTTGGATAGAGCAACAACCTTCTAAGTTGAAGGTCACAGGTTCAAATCCTGTAGAATGCG
```

```
Sequence : R|tcg|Marchantia_polymorpha
Bitscore : 95
E-value : 5.865030631201869e-23
E-value : 7.331288289002336e-24
Input : [0:73]
DB : [0:73]
Align input : GCATTCTTAGCTCAGCTGGATAGAGCAACAACCTTCTAAGTTGAAGGTCACAGGTTCAAATCCTGTAGGATGCT
Align db : GCATTCTTAGCTCAGTTGGATAGAGCAACAACCTTCTAAGTTGAAGGTCACAGGTTCAAATCCTGTAGAATGCG
```

```
Sequence : R|tcg|Marchantia_polymorpha
Bitscore : 95
E-value : 7.331288289002336e-24
Input : [0:73]
DB : [0:73]
Align input : GCATTCTTAGCTCAGCTGGATAGAGCAACAACCTTCTAAGTTGAAGGTCACAGGTTCAAATCCTGTAGGATGCT
E-value : 7.331288289002336e-24
Input : [0:73]
DB : [0:73]
E-value : 7.331288289002336e-24
Input : [0:73]
E-value : 7.331288289002336e-24
E-value : 7.331288289002336e-24
E-value : 7.331288289002336e-24
Input : [0:73]
DB : [0:73]
Align input : GCATTCTTAGCTCAGCTGGATAGAGCAACAACCTTCTAAGTTGAAGGTCACAGGTTCAAATCCTGTAGGATGCT
Align db : GCATTCTTAGCTCAGTTGGATAGAGCAACAACCTTCTAAGTTGAAGGTCACAGGTTCAAATCCTGTAGAATGCG
```

Sequence : R|tcg|Marchantia_polymorpha

Bitscore : 95

E-value : 5.865030631201869e-23

Input : [0:73]

DB : [0:73]

Align input : GCATTCTTAGCTCAGCTGGATAGAGCAACAACCTTCTAAGTTGAAGGTCACAGGTTCAAATCCTGTAGGATGCT

Align db : GCATTCTTAGCTCAGTTGGATAGAGCAACAACCTTCTAAGTTGAAGGTCACAGGTTCAAATCCTGTAGGATGCG

Sequence : R|tct|Mesostigma_viride

Bitscore : 48

E-value : 8.25429680162415e-09

Input : [30:67]

DB : [30:67]

Align input : ACCTTCTAAGTTGAAGGTCACAGGTTCAAATCCTGTAG

Align db : GCCTTCTAAGCTGTAGGTCACAGGTTCAAATCCTGTAG

Question 2 :

Unknown 1 : X|???|Malus_domestica

Meilleur hit :

M|cat|Carica_papaya (bitscore 73)

Acide aminé = Méthionine

Anticodon = CAT

Conclusion : cette séquence est un tRNA-Met (anticodon CAT).

Unknown 2: X|???|Nephroselmis_olivacea

Meilleur hit:

R|tcg|Marchantia_polymorpha (bitscore 85)

Acide aminé = Arginine

Anticodon = TCG

Conclusion : tRNA-Arg (anticodon TCG).

Unknown 3 : X|???|Phoenix_dactylifera

Meilleur hit :

P|tgg|Oryza_sativa_Japonica_Group (bitscore 40)

Acide aminé = Proline

Anticodon = TGG

Conclusion: tRNA-Pro (anticodon TGG).

Unknown 4 : X|???|Chara_vulgaris

Meilleur hit :

R|tct|Marchantia_polymorpha (bitscore 98)

Acide aminé = Arginine

Anticodon = TCT

Conclusion : tRNA-Arg (anticodon TCT).

Question 3 :

unknown 1 : X|???|Malus_domestica

PLAST :

Meilleur hit : Carica papaya tRNA-Met

Acide aminé = Méthionine (M)

Anticodon = CAT

BLASTN :

Meilleurs hits = Malus domestica, Malus sylvestris, Crataegus

Identité : 100 %

E-value : 1e-28

Annoté comme tRNA-Met

Conclusion :

PLAST et BLAST donnent la même identité fonctionnelle : tRNA-Méthionine (CAT).

BLAST précise l'espèce : Malus domestica.

unknown 2 : X|???|Nephroselmis_olivacea

PLAST :

Meilleur hit : Marchantia polymorpha

Acide aminé = Arginine (R)

Anticodon = TCG

BLASTN :

Premier hit : Nephroselmis olivacea mitochondrion

Identité : 100 %

E-value : 1e-28

Région annotée tRNA-Arg (TCG)

Conclusion :

PLAST identifie bien le type de tRNA (tRNA-Arg).

BLAST confirme l'espèce exacte : Nephroselmis olivacea.

Les deux résultats sont cohérents.

unknown 3 : X|???|Phoenix_dactylifera

PLAST :

• Meilleur hit : Oryza sativa tRNA-Pro

• Acide aminé = Proline (P)

• Anticodon = TGG

BLASTN :

• Meilleurs hits = Phoenix dactylifera (plusieurs mitochondries)

• Identité : 100 % (74/74)

• E-value : 1e-28

• Couverture : 100 %

• Annoté comme tRNA-Pro (anticodon TGG)

Conclusion :

PLAST identifie correctement la nature fonctionnelle de la séquence (tRNA-Pro, anticodon TGG).

BLAST confirme cette identification et précise l'espèce exacte, Phoenix dactylifera.

unknown 4 : X|???|Chara_vulgaris

PLAST :

• Meilleur hit : Marchantia polymorpha

• Acide aminé = Arginine (R)

• Anticodon = TCT

• Plusieurs HSPs très forts, couvrant presque toute la séquence

• PLAST identifie un tRNA-Arg

BLASTN :

• Meilleurs hits = Chara braunii, Nitellopsis obtusa, Chara vulgaris

• Identité : 100 % (74/74)

• Couverture : 100 %

• E-value : 1e-28

• Alignements mitochondriaux annotés comme tRNA-Arg (anticodon TCT)

HVX2Z594014-Alignment

Conclusion :

PLAST identifie correctement la fonction moléculaire de la séquence : un tRNA-Arg (anticodon TCT).

BLASTN confirme cette annotation et précise l'espèce d'origine : Chara vulgaris, avec une identité parfaite de 100 %.

Question 4 :

Graines plus longues

- Vitesse : plus rapide.

Une graine longue apparaît moins souvent dans la séquence, donc PLAST génère moins d'HSPs initiaux et fait moins d'extensions.

- Précision : plus élevée.

Moins de faux positifs, car les matchs de 13-15 bases consécutives sont très peu probables par hasard.

- Sensibilité : plus faible.

Une seule mutation dans la graine empêche la détection de l'HSP. PLAST manque alors plusieurs alignements biologiquement réels.

Graines plus courtes

- Vitesse : plus lente.

Beaucoup plus de k-mers correspondent dans la base de données, donc PLAST passe plus de temps à étendre des HSPs.

- Précision : plus faible.

Les graines courtes génèrent davantage de faux positifs.

- Sensibilité : plus élevée.

Les petites graines détectent plus d'alignements, même lorsque plusieurs mutations sont présentes.

Donc les graines longues favorisent la précision et la rapidité, mais réduisent la sensibilité tandis que les graines courtes favorisent la sensibilité, mais nuisent à la vitesse et à la précision.

Question 5 :

PatternHunter utilise des graines avec des positions "don't care", par exemple :

« 111010010100110111 »

Dans ce motif, seules les positions "1" doivent correspondre, ce qui permet d'accepter certaines mutations à l'intérieur même de la graine.

Adaptation de l'algorithme

Notre fonction `extract_kmers()` gère déjà naturellement les graines espacées :

- `seed[j] = "1"` // position obligatoire
- `seed[j] = "0"` // don't care

Aucune modification majeure n'est nécessaire : il suffit d'utiliser une graine espacée comme ci-dessus.

Nous avons ensuite relancé PLAST avec un seuil moins strict ($ss = 0.01$) comme demandé.

Comparaison avec la graine classique

1. Vitesse : Légèrement plus lente.

La graine espacée détecte plus de k-mers, donc PLAST étend plus d'HSPs.

2. Précision : Comparable à la graine classique.

Les positions "1" imposent suffisamment de contraintes pour éviter les faux positifs.

3. Sensibilité : Nettement plus élevée.

Grâce aux positions "0", la graine tolère des mutations internes.

PLAST trouve davantage d'alignements faibles, que la graine continue de taille 11 ne détectait pas.

Conclusion

L'utilisation d'une graine espacée de PatternHunter augmente la sensibilité de PLAST, au prix d'un léger ralentissement. Avec un seuil de significativité moins strict ($ss = 0.01$), l'algorithme détecte plus de similarités biologiques, incluant certains alignements faibles manqués par la graine classique.