

IFT 3295 Bio-informatique TP3

Yuchen Hui 20150470, Yan Zhuang 20146367

5 décembre 2022

Les codes pour ce TP peuvent se trouver dans notre [repo de Git](#). Les codes ont aussi été soumises en zip sur Studium pour une consultation plus facile.

Pour les instructions d'exécution, veuillez consulter le [README.md](#) dans le repo sus-mentionné.

1 Mise en pratique

1.1 1)

Le résultat des exécutions sur toutes les 4 séquences inconnues avec les paramètres par défaut (-ss = 0.001, -E = 4, seed = '1111111111') sont stockés dans le fichier [pratique1.txt](#) dans le github où se trouve aussi le code.

1.2 2)

Info : Le résultat pour chaque seuil se trouve dans github. [pratique1.txt](#)(-ss = 0.001) [resultat0.01.txt](#)(-ss = 0.01) [resultat0.1.txt](#)(-ss = 0.1)

1. Pour la première séquence, en jouant avec le seuil(0.001,0.01,0.1), on a trouvé que, pour n'importe quel seuil entre les trois, on obtient des séquences qui sont de type **M|cat**. Donc c'est la nature de cette séquence.
2. Pour la deuxième séquence, même si on a laissé un seuil de 0.1, c'est toujours impossible de déduire la nature de séquence, parce qu'on a plusieurs résultats qui se différent à l'un et l'autre. Toutefois, vu la similitude entre les premiers résultats, c'est peut-être possible de dire que la séquence est bien pour l'acide amine de **R**, mais c'est impossible d'être certain pour l'anticodon.
3. Pour la troisième séquence, les trois premiers résultats ont un haut score de similarité, et ils sont tous de format **P|tgg**. Donc, c'est possible de déduire que la nature de cette séquence est bien **P|tgg**
4. Pour la dernière séquence, similaire à la troisième, c'est impossible de déduire au complet la nature de séquence. Toutefois, les deux premiers résultats sont assez similaire (le score de similarité), et ils sont de format **R|**. Donc c'est possible de dire qu'elles sont de format **R|**, mais pour l'anticodon, c'est pas possible de déduire.

1.3 3)

Hypothèse : On compare les résultats générés par notre programme et BLAST avec les paramètres par défaut respectif.

Comme mentionné en haut, le résultat avec les paramètres par défaut se trouve dans le fichier [pratique1.txt](#)(-ss = 0.001) sur Github.

1. Notre seul résultat de notre programme est **M|cat|Bracteacoccus_minor_2**. Toutefois, parmi les résultats donnés BLAST, cette espèce ne figure même pas dans la liste de 100 résultats.
Les résultats de BLAST sont trop nombreux pour être inclus ici comme une capture d'écran.
2. Pour la deuxième séquence, selon les résultats de BLAST, les seuls résultats dans notre programme qui figure dans la liste de BLAST, sont ceux de l'espèce *Marchantia polymorpha*. Elles se trouvent au milieu des 20 résultats trouvés par BLAST, avec des matches de 95%

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
✓	Nephroselmis olivacea mitochondrion, complete genome	Nephroselmis olivacea	137	137	100%	1e-28	100.00%	45223	AF110138.1
✓	Candidatus Stammera capleta isolate 1 chromosome, complete genome	Candidatus Stammera...	106	106	85%	4e-19	96.83%	260485	CP043975.1
✓	Riccia fluitans strain IB-50004 mitochondrion, complete genome	Riccia fluitans	99.0	99.0	95%	6e-17	91.55%	185640	MN927134.1
✓	Riccia fluitans mitochondrion, complete genome	Riccia fluitans	99.0	99.0	95%	6e-17	91.55%	185621	NC_043906.1
✓	Riccia fluitans mitochondrion, complete genome	Riccia fluitans	99.0	99.0	95%	6e-17	91.55%	185621	MK749459.1
✓	Marchantia polymorpha subsp. ruderalis strain KBDI00084 mitochondrion, complete genome	Marchantia polymorph...	99.0	99.0	95%	6e-17	91.55%	186196	MK202951.1
✓	Marchantia polymorpha subsp. ruderalis mitochondrial DNA, complete genome, strain: Kitashirakawa-2	Marchantia polymorph...	99.0	99.0	95%	6e-17	91.55%	186196	NC_037508.1
✓	Marchantia polymorpha subsp. ruderalis Tak-1 mitochondrial DNA, complete genome	Marchantia polymorph...	99.0	99.0	95%	6e-17	91.55%	186197	AP025456.1
✓	Wiesnerella denudata mitochondrion, complete genome	Wiesnerella denudata	99.0	99.0	95%	6e-17	91.55%	186911	NC_053538.1
✓	Marchantia paleacea isolate A 18 mitochondrion, complete genome	Marchantia paleacea	99.0	99.0	95%	6e-17	91.55%	186609	M68929.1
✓	Bacillus sp. Cs-700 chromosome, complete genome	Bacillus sp. Cs-700	86.1	86.1	82%	5e-13	91.80%	4297839	CP041063.1
✓	Bacillus sp. N1-1 chromosome, complete genome	Bacillus sp. N1-1	86.1	86.1	82%	5e-13	91.80%	4497340	CP046564.1
✓	Alicyclobacillus curvatus strain ALEF1 chromosome, complete genome	Alicyclobacillus curvatus	71.3	71.3	75%	1e-08	89.29%	5970476	CP071184.1
✓	Streptobacillus moniliformis strain FDAARGOS_310 chromosome, complete genome	Streptobacillus monilifo...	69.4	69.4	67%	5e-08	92.16%	1675227	CP027400.1
✓	Streptobacillus moniliformis strain BK14232 chromosome, complete genome	Streptobacillus monilifo...	69.4	69.4	58%	5e-08	95.45%	1655770	CP079714.1
✓	Streptobacillus moniliformis DSM 12112, complete genome	Streptobacillus monilifo...	69.4	69.4	78%	5e-08	88.52%	1662578	CP001779.1
✓	Pontibacillus sp. HMF3514 chromosome, complete genome	Pontibacillus sp. HMF3...	67.6	67.6	54%	2e-07	97.50%	3976892	CP047393.1
✓	Marthasterias glacialis genome assembly, chromosome: 9	Marthasterias glacialis	56.5	56.5	78%	4e-04	84.75%	25218880	OU452227.1
✓	TPA: CrAss-like virus sp. isolate cHGM6, partial genome	CrAss-like virus sp.	52.8	52.8	37%	0.005	100.00%	85939	BK035401.1

- Pour la troisième séquence, aucun résultat donné par notre programme figure dans la liste de 100 résultats donnée par BLAST. Les résultats de BLAST sont trop nombreux pour être inclus ici comme une capture d'écran.
- Pour la dernière séquence, seulement les deux séquences qui appartiennent à l'espèce *Marchantia Polymorpha* figure dans la liste de 100 résultats de BLAST. Toutefois, elle se positionne presque à la fin de ces 100 résultats avec des matches de 97.26%

1.4 4)

Lorsque l'on utilise les graines plus longues, l'exécution sera plus rapide (car il y a moins de k-mers à considérer), la précision des résultats accroîtra (moins de séquences sont sélectionnées) et la sensibilité de l'algorithme diminue. (Avec une observation directe en déroulant l'algorithme)

Quand on utilise les graines plus courtes, l'effet sur la vitesse, précision et sensibilité est totalement inverse.