

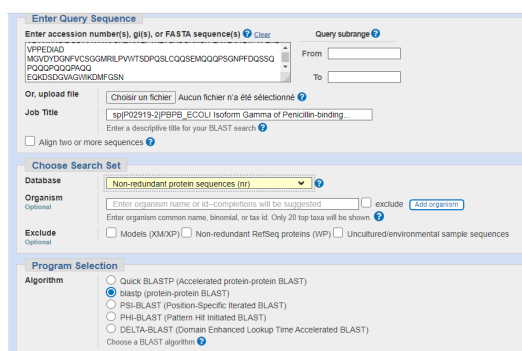
ALIGNEMENT DES SÉQUENCES PAR FASTA ET BLAST.

NOV 25, 2021

"TP - 02 BIO-INFO "

Le but de ce TP est une initiation aux outils BLAST et FASTA et à l'alignement des séquences protéiques par ces outils.

Dans la première question de TP est demandé d'identifier toutes les séquences de la protéine de la pénicilline. Ensuite on vas choisir un type de pénicilline et récupérer la séquence au format FASTA , le type choisis est : **Penicillin-binding protein 1A.**



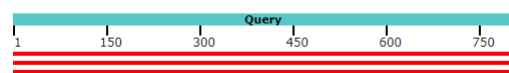
En suite, on vas faire un BLAST sur la séquence au format FASTA au NCBI, le BLAST a fait une comparaison entre la requête (QUERY) et les autres séquences (SBJECT) de la base de données qui se trouve dans NCBI , le résultat obtenu est des séquences similaire a notre requête d'un

Description	Scientific Name
✓ bifunctional glycosyl transferase/transpeptidase [Escherichia coli]	Escherichia coli
✓ TPA_bifunctional glycosyl transferase/transpeptidase [Escherichia coli]	Escherichia coli
✓ bifunctional glycosyl transferase/transpeptidase [Escherichia coli]	Escherichia coli
✓ bifunctional glycosyl transferase/transpeptidase [Escherichia coli]	Escherichia coli
✓ TPA_bifunctional glycosyl transferase/transpeptidase [Escherichia coli]	Escherichia coli
✓ TPA_bifunctional glycosyl transferase/transpeptidase [Escherichia coli]	Escherichia coli
✓ TPA_bifunctional glycosyl transferase/transpeptidase [Escherichia coli]	Escherichia coli

Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Escherichia coli	1637	1637	100%	0.0	100.00%	799	WP_211180484.1
Escherichia coli	1635	1635	100%	0.0	99.87%	827	HB05437651.1
Escherichia coli	1634	1634	100%	0.0	100.00%	847	FHM9026553.1
Escherichia coli	1634	1634	100%	0.0	100.00%	840	EFH7547147.1
Escherichia coli	1634	1634	100%	0.0	100.00%	844	HAX2053711.1
Escherichia coli	1634	1634	100%	0.0	99.87%	843	HAI7354673.1
Escherichia coli	1634	1634	100%	0.0	100.00%	844	HAW0684342.1

pourcentage varis entre 99% et 100% avec un E-value de 0.0 et Cover de 100% le cover signifit que notre requête a couvrit tout autre séquence, on peut voir aussi la version graphique de notre requête et les autre séquences, la couleur rouge veut dire que les séquences sont identique.

Distribution of the top 100 Blast Hits on 100 subject sequen



A partir de l'alignement, on peut visualiser la similarité de chaque séquences avec la notre, là ou il y'a un vide cela veut dire que les deux ne sont pas identique .

Score	Expect	Method	Identities	Positives	Gaps
1637 bits(4238)	0.0	Compositional matrix adjust.	799/799(100%)	799/799(100%)	0/79
Query 1	MPKGGKGGKGRKPRGKRGHLMLLLKLAIVFAVLIAIYGVYLDQKIRSRIDGKVNQLPAA				60
Sbjct 1	MPKGGKGGKGRKPRGKRGHLMLLLKLAIVFAVLIAIYGVYLDQKIRSRIDGKVNQLPAA				60

ALIGNEMENT DES SÉQUENCES PAR FASTA ET BLAST.

NOV 25, 2021

"TP - 02 BIO-INFO "

On va interroger la banque de données SwissProt en recherchant les séquences voisines.

On fait un BLAST à l'aide de la même séquence au format FASTA, voici une partie de résultat obtenu :

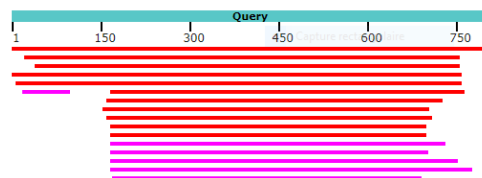
Description							
<input checked="" type="checkbox"/>	RecName: Full=Penicillin-binding protein 1B; Short=PBP-1b; Short=PBP1b; AltName: Full=Murein polymérase...						
<input checked="" type="checkbox"/>	RecName: Full=Penicillin-binding protein 1B; Short=PBP-1b; Short=PBP1b; AltName: Full=Murein polymérase...						
<input checked="" type="checkbox"/>	RecName: Full=Penicillin-binding protein 1B; Short=PBP-1b; Short=PBP1b; AltName: Full=Murein polymérase...						
<input checked="" type="checkbox"/>	RecName: Full=Penicillin-binding protein 1B; Short=PBP-1b; Short=PBP1b; AltName: Full=Murein polymérase...						
<input checked="" type="checkbox"/>	RecName: Full=Penicillin-binding protein 1B; Short=PBP-1b; Short=PBP1b; AltName: Full=Murein polymérase...						

Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Escherichia coli	1633	1633	100%	0.0	100.00%	844	P02919.2
Buchnera aphidii	868	868	91%	0.0	53.83%	760	P57296.1
Buchnera aphidii	840	840	88%	0.0	53.15%	741	Q89AR2.1
Vibrio cholerae	720	720	94%	0.0	49.74%	777	Q9KUC0.1
Haemophilus influenzae	704	704	93%	0.0	47.32%	781	P45345.1

pourcentage varis entre 26% et 100% avec un Cover entre 18 ET 100%, et différent E-value.

Le graphique de notre requête et les autres séquences, contient la couleur rouge veut dire que les séquences sont identiques, rose moins identique, le bleu et le vert montre une faible identité.

Distribution of the top 119 Blast Hits on 100 subject sequences



On remarque dans cette partie l'alignement d'un morceau de notre requête et une autre séquence, ces deux séquences ne sont pas 100% identiques.

Score	Expect	Method	Identities	Positives	Gaps
868 bits(2243)	0.0	Compositional matrix adjust.	394/732(54%)	542/732(74%)	5/732(0%)
Query 22	WLLKLAIVFAVLIAIYGVYVLDQKTRSDGKQVQLPAAYVGRVWILEPDMTISKNEVK	81	+ L+K+ +L YG+YL KI I+GKVN P ++YGR+VILEP + S+ E++		
Subjct 8	YFLIKVFFVLILTLCTGLYLYVKINRFINGKVNFPSTIYGRIVNLEPGNSYSQKEVLH	67			

Conclusion

les séquences trouvées ne sont pas les mêmes que celle trouvées par BLAST.

HANIFI FATIMA EL-BATOUL
GROUPE - 02 - IAA