



Tutoriel d'utilisation du logiciel BLAST

BLAST (« Basic Local Alignment Search Tool ») est un logiciel développé par le National Center for Biotechnology Information qui permet de :

- Comparer statistiquement une séquence de nucléotides ou de protéines sélectionnées, à des bases de données comprenant des milliers de séquences connues.
- Déterminer l'identité de cette séquence (« de quelle séquence ma séquence est la plus proche ») : à quelle espèce elle appartient, et quelle est la fonction de cette séquence.

Sommaire

1. Récupérer les séquences d'intérêt	1
2. Accéder à BLAST	1
3. Comparer la séquence à la base de données	2
4. Comprendre les résultats	3
5. Ajouter l'espèce dans le tableur numérique	3

1. Récupérer les séquences d'intérêt

Dans le dossier du kit, le tableur numérique « sequences-adn » contient des séquences ADN inconnues dans l'onglet « séquences_élèves ».

Sélectionner une séquence dans la colonne "séquences d'ADN à identifier", et la copier.

2. Accéder à BLAST

a. Cliquer sur le lien : <https://blast.ncbi.nlm.nih.gov/Blast.cgi> ou copier le lien dans votre navigateur internet.

b. Cliquer sur « Nucleotide BLAST » (encadré rouge sur l'image ci-contre).

Cet outil permet de comparer une séquence de nucléotides à d'autres séquences de nucléotides.



3. Comparer la séquence à la base de données

a. Dans le champ « Enter Query Sequence », coller la séquence sélectionnée. Cocher ensuite « Show results in a new window » et cliquer sur « Algorithm parameters » (voir encadrés rouges sur l'image ci-dessous).

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#)

Query subrange [?](#)

From

To

Or, upload file Aucun fichier sélectionné. [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Choose Search Set

Database Standard databases (nr etc.): rRNA/ITS databases Genomic + transcript databases Betacoronavirus

Nucleotide collection (nr/nt) [?](#)

Organism exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude Models (XM/XP) Uncultured/environmental sample sequences

Limit to Sequences from type material

Entrez Query [YouTube](#) [Create custom database](#)

Enter an Entrez query to limit search [?](#)

Program Selection

Optimize for Highly similar sequences (megablast) More dissimilar sequences (discontiguous megablast) Somewhat similar sequences (blastn)

Choose a BLAST algorithm [?](#)

Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)

Show results in a new window

b. Dans le champ « Max matches in a query range » (encadré rouge sur l'image ci-dessous), rentrer 1. Cela permettra d'afficher uniquement la meilleure correspondance de la séquence copiée.

Cliquer ensuite sur le bouton « BLAST ».

Le serveur va chercher la meilleure correspondance de la séquence (cela peut prendre plusieurs secondes).

General Parameters

Max target sequences [?](#)

Select the maximum number of aligned sequences to display [?](#)

Short queries Automatically adjust parameters for short input sequences [?](#)

Expect threshold [?](#)

Word size [?](#)

Max matches in a query range [?](#)

Scoring Parameters

Match/Mismatch Scores [?](#)

Gap Costs [?](#)

Filters and Masking

Filter Low complexity regions [?](#) Species-specific repeats for: [?](#)

Mask Mask for lookup table only [?](#) Mask lower case letters [?](#)

Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)

Show results in a new window

4. Comprendre les résultats

La fenêtre de résultats apparaît et donne plusieurs indications :

- Le type de séquence « Molecule type ». Dans l'exemple ci-dessous, la molécule est de type « dna », c'est-à-dire qu'il s'agit d'une séquence nucléotidique.
- La longueur de la séquence « query length ». Ici, elle mesure 88 nucléotides.
- Le nom de l'espèce qui a le plus de correspondance avec la séquence. Ici il s'agit d'« Emilia huxleyi ».

Job Title: Nucleotide Sequence
RID: KAM4TED6013
Program: BLASTN
Database: nt
Query ID: lcl|Query_13441
Description: None
Molecule type: dna
Query Length: 88
Other reports: Distance tree of results, MSA viewer

Filter Results
Organism: only top 20 will appear
Percent Identity: [] to []
E value: [] to []
Query Coverage: [] to []
Filter, Reset

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments
Download | Select columns | Show 100

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Emiliana huxleyi CCMP1516 hypothetical protein partial mRNA	Emiliana huxleyi CCMP1516	135	135	100%	6e-28	94.32%	1327	XM_005756271.1

5. Ajouter l'espèce dans le tableur numérique

Copier le nom de l'espèce et le coller dans le tableur numérique « sequences-adn », dans la colonne « nom de l'organisme probable », à côté de la séquence ADN associée. Les corrections se trouvent dans l'onglet « correctio_séquences_identifiées ».