

1 MESURE DE LA BIODIVERSITÉ À L'AIDE DE BLAST

Complément numérique • Protocole d'utilisation de BLAST
Manuel p. 188-189 – chapitre 9

BLAST (*Basic Local Alignment Search Tool*, du *National Center for Biotechnology Information* – NCBI, E-U) est un algorithme qui permet :

- de comparer une séquence d'intérêt à une banque de données constituée de milliards de séquences
- de 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.

1. Récupérer les séquences d'intérêt

Dans le fichier EST_c09_act1_sequences.xls, dans la feuille « Séquences », copier la séquence A en gardant l'en-tête >sequence-A et en faisant bien attention de ne sélectionner que le texte, et non la cellule entière.

2. Accéder à l'outil BLAST

a. Cliquer sur le lien : <https://blast.ncbi.nlm.nih.gov/Blast.cgi> ou entrer « Blast » dans un moteur de recherche et cliquer sur le premier lien.

Remarque : il existe également d'autres utilisations de cet algorithme, en utilisant des séquences d'acides aminés. Nous ne les utiliserons pas ici.

b. Cliquer sur **Nucleotide BLAST** : 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**, copier la séquence d'intérêt (par exemple, la séquence A du fichier). Cocher ensuite **Show results in a new window** et cliquer sur **Algorithm parameters**.

Standard Nucleotide BLAST

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

Enter Query Sequence

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

From

To

Or, upload file

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
Optional exclude
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

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

Limit to Sequences from type material
Optional

Entrez Query
Optional [You \[this\]\(#\) \[Create custom database\]\(#\)](#)
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

[+ Algorithm parameters](#) [Note: Parameter values that differ from the default are highlighted in yellow and marked with + sign](#)

b. Dans le champ **Max matches in a query range**, rentrer 1. Cela permettra d'afficher uniquement la meilleure correspondance de la séquence copiée. Cliquer ensuite sur le bouton bleu **BLAST**. Le serveur va chercher la meilleure correspondance de la séquence. Cela peut prendre quelques secondes à quelques dizaines de secondes selon l'utilisation du serveur (ce site est utilisé en temps réel par des chercheurs du monde entier).

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

Show results in a new window

[+ Algorithm parameters](#) [Note: Parameter values that differ from the default are highlighted in yellow and marked with + sign](#)

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 indique le nom du « job » (ici, sequence-A). La molécule est de type « dna » (séquence nucléotidique), et elle mesure 648 nucléotides.

Job Title	sequence-A
RID	NHXK414S016 <small>Search expires on 09-10 23:36 pm</small> Download All ▼
Program	BLASTN Citation ▼
Database	nt See details ▼
Query ID	lc Query_28045
Description	sequence-A
Molecule type	dna
Query Length	648
Other reports	Distance tree of results MSA viewer ?

Filter Results

Organism only top 20 will appear exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity to **E value** to **Query Coverage** to

[Filter](#) [Reset](#)

Descriptions		Graphic Summary	Alignments	Taxonomy			
Sequences producing significant alignments Download ▼ Manage Columns ▼ Show <input type="text" value="100"/> ?							
<input checked="" type="checkbox"/> select all <small>1 sequences selected</small>		GenBank Graphics Distance tree of results					
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	Emiliana huxleyi cytochrome b6 (petB) gene, complete cds; chloroplast	1197	1197	100%	0.0	100.00%	AY675518.1

5. Ajouter l'espèce dans le tableau Excel

Dans le tableau de l'autre feuille du fichier EST_c09_act1_sequences.xls, en face de la séquence A, ajouter le nom d'espèce : *Emiliana huxleyi*.

Se renseigner, en utilisant Internet, sur le groupe auquel appartient cet organisme.