

mgr Katarzyna Stefańska

Katedra i Zakład Informatyki i Statystyki

Zalecenia dotyczące przyrównywania sekwencji nukleotydowych

Wykorzystując BLAST

data przygotowania: 2024-04-22

ŚCIEŻKI DOSTĘPU

Podstawa teoretyczna

Narzędzie BLAST (Basic Local Alignment Search Tool) jest dostępne na stronie internetowej NCBI (National Center for Biotechnology Information) i służy do przyrównywania sekwencji nukleotydowych i aminokwasowych.

Krok 1: Wczytaj sekwencję w programie BLAST

1. Otwórz dowolną przeglądarkę internetową na swoim urządzeniu.
2. Przejdź do strony BLAST: w pasku adresu przeglądarki wpisz adres strony <https://blast.ncbi.nlm.nih.gov/Blast.cgi>.
3. Naciśnij klawisz Enter, aby przejść do strony głównej BLAST.
4. Na stronie głównej BLAST znajdują się różne programy BLAST, takie jak BLAST Nucleotide (do przyrównywania sekwencji nukleotydowych), BLAST Protein (do przyrównywania sekwencji białkowych) itp.
5. Wybierz program **Nucleotide BLAST**, jeśli chcesz przyrównać sekwencje nukleotydowe.

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

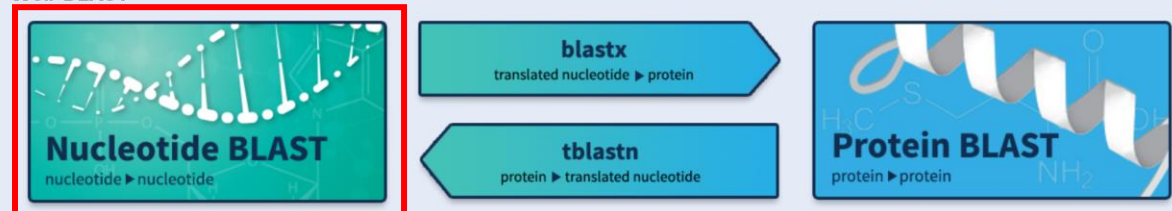
BLAST+ 2.15.0 is here!

We have included two exciting new features in the latest BLAST+ release

Tue, 28 Nov 2023

[More BLAST news...](#)

Web BLAST



The screenshot shows the Web BLAST interface. On the left, the 'Nucleotide BLAST' option is highlighted with a red box. In the center, there are two buttons: 'blastx translated nucleotide → protein' and 'tblastn protein → translated nucleotide'. On the right, the 'Protein BLAST protein → protein' option is visible. The background features a DNA double helix and a protein ribbon structure.

6. Na stronie Nucleotide BLAST znajduje się pole do wprowadzania sekwencji – możesz wczytać sekwencję poprzez wklejenie jej bezpośrednio do pola tekstowego, zaimportować plik zawierający sekwencję lub wpisać numer dostępu („accession number”) konkretnej sekwencji z GenBanku.

Krok 2: Dostosuj ustawienia wyszukiwania i uruchom wyszukiwanie

1. Ustaw zakres przeszukiwania sekwencji w polu **Query subrange**. Jeśli chcesz do przyrównania wykorzystać całą wklejoną sekwencję, pozostaw pole **Query subrange puste**.
2. Wybierz bazę danych (**Database**), która zostanie wykorzystana do przeszukiwania – w przypadku, gdy nie masz wielu informacji o sekwencji wybierz **Standard databases** oraz **Nucleotide collection** dla sekwencji nukleotydowych.
3. Opcjonalnie możesz wybrać organizm (**Organism**), z którego pochodzi sekwencja wykorzystana do przyrównania.
4. Jeśli nie masz konkretnych wymagań co do przyrównania, pozostałe opcje zostaw domyślnie.
5. Kliknij na przycisk **BLAST**, aby uruchomić program.

BLASTN programs search nucleotide databases using a

Enter Query Sequence

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

From

To

Or, upload file Nie wybrano pliku. [?](#)

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

[New](#) Experimental databases [Try experimental taxonomic nt databases](#) [Download](#)

For more info see [What are taxonomic nt databases?](#)

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

Enter organism name or id—completions will be suggested exclude [Add organism](#)

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

Models (XM/XP) Uncultured/environmental sample sequences

Sequences from type material

Limit to

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 [?](#)

BLAST Search database nt using Megablast (Optimize for highly similar sequences)

Show results in a new window

[+ Algorithm parameters](#)

wklejamy „accession number”
albo sekwencję nukleotydów

wyberamy bazę danych

wyberamy organizm

Krok 3: Analiza wyników

- Po zakończeniu procesu wyszukiwania zostaną wyświetlone wyniki przyrównania sekwencji.
- Wyniki BLAST są zwykle prezentowane w formie tabelarycznej lub graficznej, co ułatwia analizę podobieństw między sekwencjami. Dla każdego trafienia BLAST podaje szczegółowe informacje na temat przyrównania, w tym podobieństwo, E-value, identyczność, pokrycie sekwencji.

Job Title	NM_006563:Homo sapiens KLF transcription factor...
RID	2CWJK2V2016 <small>Search expires on 04-23 22:21 pm</small> Download All ▾
Program	BLASTN Citation ▾
Database	nt See details ▾
Query ID	lcl Query_3641769
Description	None
Molecule type	dna
Query Length	1400
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

Download ▾ Select columns ▾ Show 100 ▾ [?](#)

Sequences producing significant alignments

select all 100 sequences selected [GenBank](#) [Graphics](#) [Distance tree of results](#) [MSA Viewer](#)

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Homo sapiens KLF transcription factor 1 (KLF1), mRNA	Homo sapiens	2586	2586	100%	0.0	100.00%	1613	NM_006563.5
<input checked="" type="checkbox"/> Homo sapiens Kruppel-like factor 1 (erythroid), mRNA (cDNA clone IMAGE:5243387)	Homo sapiens	2586	2586	100%	0.0	100.00%	1623	BC040000.1
<input checked="" type="checkbox"/> Homo sapiens Kruppel-like factor 1 (erythroid), mRNA (cDNA clone MGC:34237 IMAGE:5201847), com...	Homo sapiens	2580	2580	100%	0.0	99.93%	1646	BC033580.1
<input checked="" type="checkbox"/> Human erythroid-specific transcription factor EKLF mRNA, complete cds	Homo sapiens	2558	2558	100%	0.0	99.64%	1613	U65404.1
<input checked="" type="checkbox"/> PREDICTED: Pan paniscus KLF transcription factor 1 (KLF1), mRNA	Pan paniscus	2547	2547	100%	0.0	99.50%	2109	XM_003814829.3
<input checked="" type="checkbox"/> PREDICTED: Pan troglodytes KLF transcription factor 1 (KLF1), mRNA	Pan troglodytes	2542	2542	100%	0.0	99.43%	1633	XM_524128.7
<input checked="" type="checkbox"/> PREDICTED: Pongo abelii KLF transcription factor 1 (KLF1), mRNA	Pongo abelii	2431	2431	100%	0.0	98.00%	2116	XM_002828744.5
<input checked="" type="checkbox"/> PREDICTED: Nomascus leucogenys Kruppel like factor 1 (KLF1), mRNA	Nomascus leuc...	2425	2425	100%	0.0	97.93%	1575	XM_003275591.3
<input checked="" type="checkbox"/> PREDICTED: Pongo pygmaeus KLF transcription factor 1 (KLF1), mRNA	Pongo pygmaeus	2420	2420	100%	0.0	97.86%	1552	XM_054464186.1
<input checked="" type="checkbox"/> PREDICTED: Hylobates moloch KLF transcription factor 1 (KLF1), mRNA	Hylobates moloch	2407	2407	100%	0.0	97.72%	1551	XM_032169856.2
<input checked="" type="checkbox"/> PREDICTED: Symphalangus syndactylus KLF transcription factor 1 (KLF1), mRNA	Symphalangus ...	2401	2401	100%	0.0	97.64%	1551	XM_055237479.1

3. Po kliknięciu w konkretne trafienie użytkownik może zobaczyć przyrównanie sekwencji zapytania i trafienia, co pozwala na dokładne zrozumienie podobieństw i różnic między nimi.
4. Jeśli trafienie odnosi się do istniejącej sekwencji w bazie GenBank, użytkownik może uzyskać dostęp do pełnych informacji o tej sekwencji klikając w odnośnik **GenBank**.

Homo sapiens KLF transcription factor 1 (KLF1), mRNA

Sequence ID: [NM_006563.5](#) Length: 1613 Number of Matches: 1

Range 1: 71 to 1470 [GenBank](#) [Graphics](#)

[▼ Next Match ▲](#)

Score	Expect	Identities	Gaps	Strand
2586 bits(1400)	0.0	1400/1400(100%)	0/1400(0%)	Plus/Plus
Query 1		GCCGAGACCGCCTTGCCCTCCATCAGCACACTGACCGCCCTGGGCCCTTCCCGGACACA		60
Sbjct 71		GCCGAGACCGCCTTGCCCTCCATCAGCACACTGACCGCCCTGGGCCCTTCCCGGACACA		130
Query 61		CAGGATGACTTCTCAAGTGGTGGCGCTCCGAAGAGGCGCAGGACATGGGCCCGGGTCCT		120
Sbjct 131		CAGGATGACTTCTCAAGTGGTGGCGCTCCGAAGAGGCGCAGGACATGGGCCCGGGTCCT		190
Query 121		CCTGACCCACGGAGCCGCCCTCCACGTGAAGTCTGAGGACCAGCCCGGGAGGAAGAG		180
Sbjct 191		CCTGACCCACGGAGCCGCCCTCCACGTGAAGTCTGAGGACCAGCCCGGGAGGAAGAG		250
Query 181		GACGATGAGAGGGGCGCGGACGCCACCTGGGACCTGGATCTCCTCCTACCAACTTCTCG		240
Sbjct 251		GACGATGAGAGGGGCGCGGACGCCACCTGGGACCTGGATCTCCTCCTACCAACTTCTCG		310
Query 241		GGCCCGGAGCCCGGTGGCGCGCCCAAGACCTGCGCTCTGGCGCCAGCGAGGCCCTCCGGG		300
Sbjct 311		GGCCCGGAGCCCGGTGGCGCGCCCAAGACCTGCGCTCTGGCGCCAGCGAGGCCCTCCGGG		370
Query 301		GCGCAATATCCGCCGCCGCCGAGACTCTGGGCGCATATGCTGGCGGCCCGGGGCTGGTG		360
Sbjct 371		GCGCAATATCCGCCGCCGCCGAGACTCTGGGCGCATATGCTGGCGGCCCGGGGCTGGTG		430
Query 361		GCTGGGCTTTTGGGTTTCGGAGGATCACTCGGGTTGGGTGCGCCCTGCCCTGCGAGCCCGG		420
Sbjct 431		GCTGGGCTTTTGGGTTTCGGAGGATCACTCGGGTTGGGTGCGCCCTGCCCTGCGAGCCCGG		490
Query 421		GCTCCCGACGCCTTCGTGGGCCAGCCCTGGCTCCAGCCCCGGCCCCGAGCCCAAGGCG		480
Sbjct 491		GCTCCCGACGCCTTCGTGGGCCAGCCCTGGCTCCAGCCCCGGCCCCGAGCCCAAGGCG		550