



Escuela Técnica Superior de
Ingeniería Informática

Técnicas Inteligentes en Bioinformática

Tarea 4

Profesores

David Orellana Martin

Pedro Almagro Blanco

Alumno

Germán Lorenz Vieta

Esta tarea está íntimamente ligada con la clase impartida sobre BLAST.

Para ello, vamos a realizar los siguientes pasos:

Con lo que ya hemos visto para las proteínas, podemos realizar este ejercicio:

- Acceder al servidor de NCBI
- Determinar si esta secuencia 2 de ADN que hemos obtenido en un secuenciador tiene similitud con algunos genes conocidos, para poder consultar anotaciones.
- Quedarnos con la primera secuencia cuyo gen tenga RefSeq a revisad (accession NM ...).
- ¿Cómo se llama el gen? ¿Podemos en principio considerarlo homólogo?

Para poder verificar similitud con otra secuencia utilice BLAST según lo especificado en lo explicado en clase <https://www.cs.us.es/~dorellana/blast.html>

Descargue la secuencia SecuenciaADN.txt y la envíe a procesar en BLAST configurando la database como refseq_rna y la selección en blastn

The screenshot shows the NCBI BLAST web interface. The 'Enter Query Sequence' section has a text input field containing a nucleotide sequence: 'CGTGTGCTTACATCTCACAATATTTGGGGAAGCAAAAGCCAAATATAAGGATGATTTTATATTGTTTAAATTTAAATTTTAAATGTTGGAATCTTCATGTGAAAATTATATGTATGTAAATATGCATTACACTTGGAGGCAA'. The 'Choose Search Set' section has 'Database' set to 'Reference RNA sequences (refseq_rna)'. The 'Program Selection' section has 'Optimize for' set to 'Somewhat similar sequences (blastn)'. The 'BLAST' button is visible at the bottom.

El procesamiento inicio y luego de unos minutos obtuve los alineamientos más significativos

[Edit Search](#)
[Save Search](#)
[Search Summary](#)
[How to read this report?](#)
[BLAST Help Videos](#)
[Back to Traditional Results Page](#)

Job Title unknown

RID AM8HS4GK013 Search expires on 06-17 02:17 am [Download All](#)

Program BLASTN [Citation](#)

Database refseq_ma [See details](#)

Query ID IclQuery_64079

Description unknown

Molecule type dna

Query Length 11001

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

☒ 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"/> | PREDICTED: Drosophila yakuba protein BCL9 homolog (LOC6523724), transcript variant X3, mRNA | Drosophila yakuba | 3627 | 9070 | 45% | 0.0 | 100.00% | 5016 | XM_039377129.2 |
| <input checked="" type="checkbox"/> | PREDICTED: Drosophila yakuba protein BCL9 homolog (LOC6523724), transcript variant X2, mRNA | Drosophila yakuba | 3627 | 9226 | 46% | 0.0 | 100.00% | 5102 | XM_002099563.4 |
| <input checked="" type="checkbox"/> | PREDICTED: Drosophila yakuba protein BCL9 homolog (LOC6523724), transcript variant X1, mRNA | Drosophila yakuba | 3627 | 9358 | 47% | 0.0 | 100.00% | 5174 | XM_015191338.3 |
| <input checked="" type="checkbox"/> | PREDICTED: Drosophila santomea protein BCL9 homolog (LOC120454922), transcript variant X4, mRNA | Drosophila santomea | 3578 | 8895 | 45% | 0.0 | 99.45% | 5011 | XM_039640670.2 |
| <input checked="" type="checkbox"/> | PREDICTED: Drosophila santomea protein BCL9 homolog (LOC120454922), transcript variant X3, mRNA | Drosophila santomea | 3578 | 8972 | 46% | 0.0 | 99.45% | 5048 | XM_039640668.2 |
| <input checked="" type="checkbox"/> | PREDICTED: Drosophila santomea protein BCL9 homolog (LOC120454922), transcript variant X2, mRNA | Drosophila santomea | 3578 | 9051 | 46% | 0.0 | 99.45% | 5097 | XM_039640667.2 |
| <input checked="" type="checkbox"/> | PREDICTED: Drosophila santomea protein BCL9 homolog (LOC120454922), transcript variant X1, mRNA | Drosophila santomea | 3578 | 9122 | 46% | 0.0 | 99.45% | 5156 | XM_039640669.2 |

Podemos observar varias secuencias similares en un 100% asociadas con Drosophila yakuba protein BCL9 homolog

Respecto al análisis la similitud se predice Drosophila yakuba protein BCL9

[Descriptions](#) [Graphic Summary](#) **[Alignments](#)** [Taxonomy](#)

Alignment view Pairwise ☐ CDS feature [Restore defaults](#) [Download](#)

100 sequences selected

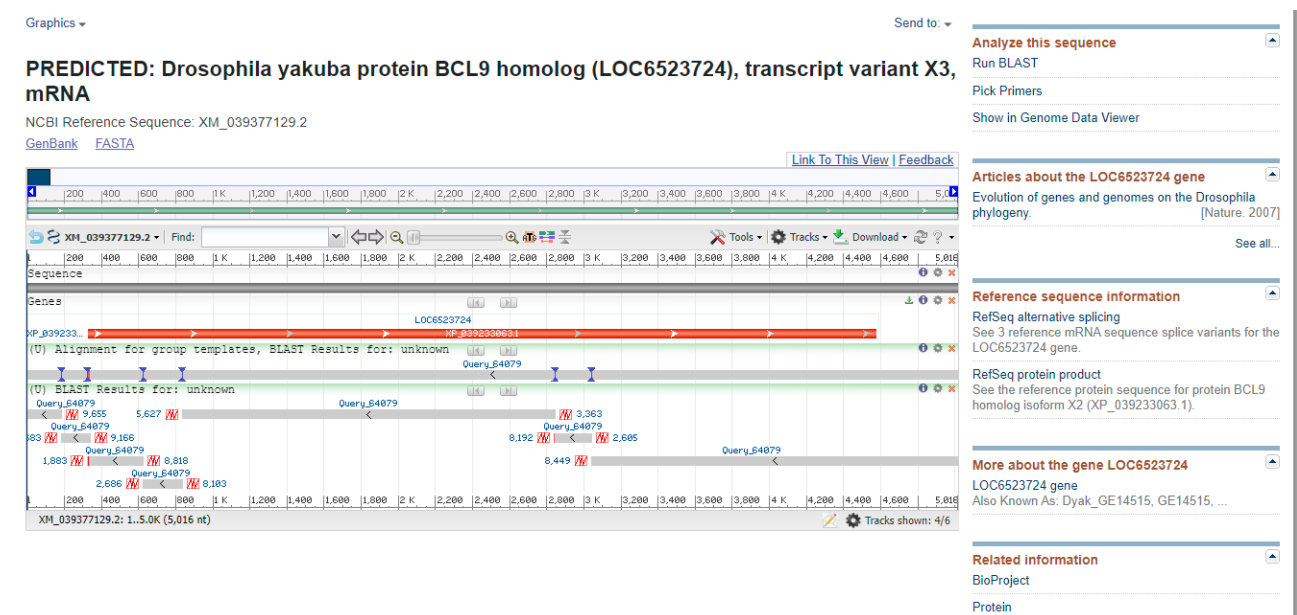
[Download](#) [GenBank](#) [Graphics](#) Sort by: E value [Next](#) [Previous](#) [Descriptions](#)

PREDICTED: Drosophila yakuba protein BCL9 homolog (LOC6523724), transcript variant X3, mRNA
Sequence ID: [XM_039377129.2](#) Length: 5016 Number of Matches: 7

Range 1: 835 to 2845 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

| Score | Expect | Identities | Gaps | Strand |
|-----------------|--|-----------------|------------|------------|
| 3627 bits(4022) | 0.0 | 2011/2011(100%) | 0/2011(0%) | Plus/Minus |
| Query 3364 | CAGCAGAGGACTGACCGAATGAGTCCAATTCCTTTGGTGATAGATGGGATAGAGGGGTAC | 3423 | | |
| Sbjct 2845 | CAGCAGAGGACTGACCGAATGAGTCCAATTCCTTTGGTGATAGATGGGATAGAGGGGTAC | 2786 | | |
| Query 3424 | TTGGGTTGCTGTTTAAGTTATGCGTAAGAAGCGCTTGATGATGCGGATTTCTACTTCGAT | 3483 | | |
| Sbjct 2785 | TTGGGTTGCTGTTTAAGTTATGCGTAAGAAGCGCTTGATGATGCGGATTTCTACTTCGAT | 2726 | | |
| Query 3484 | TTTGATTTGACGGCGATGGGGTGTCTGCTGAAAACAGTTTGTGTTGCTGAGAGCACTG | 3543 | | |
| Sbjct 2725 | TTTGATTTGACGGCGATGGGGTGTCTGCTGAAAACAGTTTGTGTTGCTGAGAGCACTG | 2666 | | |
| Query 3544 | TTGTTGTGCCAGCCTGAGCCGCCGACGTATTAGCTTGTGGAGCAGATCCAGATAATGATC | 3603 | | |
| Sbjct 2665 | TTGTTGTGCCAGCCTGAGCCGCCGACGTATTAGCTTGTGGAGCAGATCCAGATAATGATC | 2606 | | |
| Query 3604 | CTGATCCATCCATGCTGGGAGAGTTGGTTGGCAAGCCCATGACTGCTCCAGTTGTGCGG | 3663 | | |
| Sbjct 2605 | CTGATCCATCCATGCTGGGAGAGTTGGTTGGCAAGCCCATGACTGCTCCAGTTGTGCGG | 2546 | | |
| Query 3664 | GTGACGGCAGGATAGGTTGTTGGACTCGAGGATTGGGTGATTGAGTGGCTATGGGTA | 3723 | | |
| Sbjct 2545 | GTGACGGCAGGATAGGTTGTTGGACTCGAGGATTGGGTGATTGAGTGGCTATGGGTA | 2486 | | |
| Query 3724 | CTGACGCCGATCGCTGGGTTGAGTGGTAAGGAGGAGGAGACCTTGACAGGTTCCGACCT | 3783 | | |
| Sbjct 2485 | CTGACGCCGATCGCTGGGTTGAGTGGTAAGGAGGAGGAGACCTTGACAGGTTCCGACCT | 2426 | | |
| Query 3784 | GGCTGCTGACGAGTTTCCACAGTTGCAAGGAGTTTGTGTTGCGGTGCTACATTTC | 3843 | | |

De forma grafica



Ingreso por Accesion NM al primero para verificarlo

[https://www.ncbi.nlm.nih.gov/nucleotide/XM_039377129.2?report=genbank&log\\$=nuclo p&blast_rank=1&RID=AM8HS4GK013](https://www.ncbi.nlm.nih.gov/nucleotide/XM_039377129.2?report=genbank&log$=nuclo p&blast_rank=1&RID=AM8HS4GK013)

La proteina es BCL9, el gen LOC6523724 y la variante x3

Drosophila yakuba protein BCL9 homolog (LOC6523724), transcript variant X3, mRNA

The screenshot displays the NCBI GenBank record for the predicted Drosophila yakuba protein BCL9 homolog (LOC6523724), transcript variant X3, mRNA. The record includes the accession number, definition, accession, version, DBLINK, keywords, source, organism, and comment. The accession number is XM_039377129.2. The definition is PREDICTED: Drosophila yakuba protein BCL9 homolog (LOC6523724), transcript variant X3, mRNA. The accession is XM_039377129. The version is XM_039377129.2. The DBLINK is BioProject: PRJNA695779. The keywords are RefSeq. The source is Drosophila yakuba. The organism is Drosophila yakuba. The comment is MODEL REFSEQ: This record is predicted by automated computational analysis. This record is derived from a genomic sequence.

Entonces por el análisis de similitud realizado en BLAST podemos considerarlo homólogo