

Oct 31, 2025

# Procolo de análisis bioinformático elemental: desde acceso al GenBank al análisis genético

DOI

<https://dx.doi.org/10.17504/protocols.io.kxygx48zwl8j/v1>

Dumar Jaramillo-Hernández<sup>1</sup>, If.salazar<sup>2</sup>

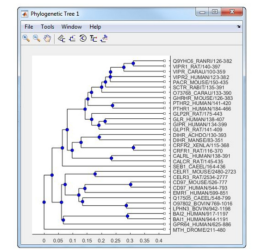
<sup>1</sup>Universidad de los Llanos (Colombia); <sup>2</sup>Universidad Técnica de Ambato (Ecuador)

Dumar\_Jaramillo\_H



Dumar A Jaramillo-Hernández

Universidad de los Llanos



## Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.kxygx48zwl8j/v1>

**Protocol Citation:** Dumar Jaramillo-Hernández, If.salazar 2025. Procolo de análisis bioinformático elemental: desde acceso al GenBank al análisis genético. **protocols.io** <https://dx.doi.org/10.17504/protocols.io.kxygx48zwl8j/v1>

**License:** This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Protocol status:** Working

**We use this protocol and it's working**

**Created:** October 29, 2025

**Last Modified:** October 31, 2025

**Protocol Integer ID:** 231080

**Keywords:** Bioinformática, Toxocara, Filogenia, análisis bioinformático básico asociado al ingreso al genbank, al genbank al análisis genético protocolo, genbank to genetic analysis, basic bioinformatics analysis protocol, análisis bioinformático, genetic analysis, proteínas de diversos organismo, genes from microorganism, sequencing gene fragment, genetic distances between genbank repository, distanciamientos genéticos entre, accessing genbank, result of molecular diagnostic product, gene repository, proteins from various organism, gene fragment, genbank repository, molecular diagnostic product, subsequent phylogenetic analysis, microorganism, comparing genetic distance, gene, genbank, procesos elementales para entender principio, various organism, complete genome

## Abstract

Protocolo de análisis bioinformático básico asociado al ingreso al GenBank, búsqueda de depósito de genes, fragmentos de genes, genomas completos o proteínas de diversos organismos, posterior principio de análisis filogenético usando secuencias FAST de los depósitos y mecanismos de comparaciones de distanciamientos genéticos entre depósitos del GenBank. Procesos elementales para entender principios de acciones a seguir posterior a la secuenciación de fragmentos génicos o genes de microorganismos, como resultado de productos de diagnósticos molecular (ej. PCR convencional).

### Basic bioinformatics analysis protocol: from accessing GenBank to genetic analysis

Basic bioinformatics analysis protocol associated with accessing GenBank, searching for gene repositories, gene fragments, complete genomes, or proteins from various organisms, subsequent phylogenetic analysis using FAST sequences from the repositories, and mechanisms for comparing genetic distances between GenBank repositories. Elementary processes for understanding the principles of actions to be taken after sequencing gene fragments or genes from microorganisms, as a result of molecular diagnostic products (e.g., conventional PCR).

## Attachments



Metodología análisis...

2.9MB

## Guidelines

Otras bases de datos abiertas de depósitos de información génica o proteoma en el mundo que deben ser consultadas

|  | Base de Datos         | Qué contiene                                                 | Enlace                                                                                    |
|--|-----------------------|--------------------------------------------------------------|-------------------------------------------------------------------------------------------|
|  | <b>NCBI GenBank</b>   | Secuencias de ADN, ARN, proteínas                            | <a href="https://www.ncbi.nlm.nih.gov/genbank/">https://www.ncbi.nlm.nih.gov/genbank/</a> |
|  | <b>EMBL-EBI (ENA)</b> | Archivo europeo de secuencias                                | <a href="https://www.ebi.ac.uk/ena">https://www.ebi.ac.uk/ena</a>                         |
|  | <b>DDBJ (Japón)</b>   | Parte del consorcio internacional con NCBI y EMBL            | <a href="https://www.ddbj.nig.ac.jp/">https://www.ddbj.nig.ac.jp/</a>                     |
|  | <b>UniProt</b>        | Secuencias y anotaciones de proteínas                        | <a href="https://www.uniprot.org/">https://www.uniprot.org/</a>                           |
|  | <b>BOLD Systems</b>   | Secuencias de código de barras genético (usado en taxonomía) | <a href="https://www.boldsystems.org/">https://www.boldsystems.org/</a>                   |

## Materials

Computador.

Se puede trabajar desde CDM de windows o desde la terminal de Mac.

Acceso a internet.

## Safety warnings

### ! MrBayes IQ-TREE y BEAST

Importante, si bien son herramientas muy robustas, se necesita tener conocimientos previos de lenguaje de programación, se puede trabajar desde CDM de windows o desde la terminal de Mac

## Ethics statement

Son experimentos in-silico, exclusivamente análisis computacionales de información génica o proteoma dispuesta en la web en bases de datos. Dependiendo de los usos que le darán a la información analizada y generada, deberán solicitar avales a los IACUC.



## Before start

### Programas para alineamiento de secuencias:

**Clustal Omega** Para alineamientos múltiples. <https://www.ebi.ac.uk/Tools/msa/clustalo/>

**MAFFT** Para grandes volúmenes de secuencias. <https://mafft.cbrc.jp/alignment/server/>

**MUSCLE** Alineamientos múltiples. <https://www.ebi.ac.uk/jdispatcher/msa/muscle?type=protein>

### Programas para construir árboles filogenéticos:

**MEGA (Molecular Evolutionary Genetics Analysis)** <https://www.megasoftware.net/>

- Permite alineamientos, construcción y visualización de árboles.
- Ideal para principiantes

**PhyML** <http://www.atgc-montpellier.fr/phyml/>

- Usa métodos de Máxima Verosimilitud para árboles más robustos.

**RAxML (Randomized Axelerated Maximum Likelihood)** <https://evomics.org/learning/phylogenetics/raxml/>

- Muy eficiente para grandes conjuntos de datos.
- Basado en máxima verosimilitud.

**MrBayes** <https://sourceforge.net/projects/mrbayes/>

- Utiliza inferencia bayesiana para construir árboles con estimaciones probabilísticas.

**IQ-TREE** <https://iqtree.github.io/>



Rápido y automatizado, muy usado en filogenias modernas.

**BEAST** <https://beast.community/>

- Especializado en análisis filogenéticos con información temporal (ej. evolución viral).

### **Visualización de árboles filogenéticos:**

**FigTree** <http://tree.bio.ed.ac.uk/software/figtree/>

- Intuitivo y útil para editar y personalizar árboles.

**iTOL (Interactive Tree Of Life)** <https://itol.embl.de/>

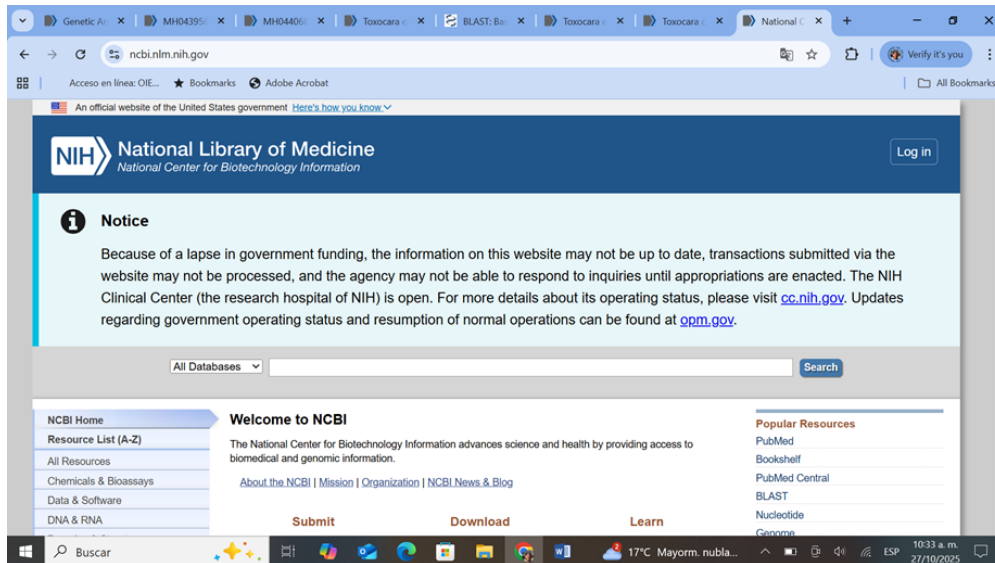
- Permite crear árboles interactivos, con colores, etiquetas, etc.

## Metodología desde el acceso al GenBank al análisis genético

- 1 Ingresa a la base datos <https://www.ncbi.nlm.nih.gov/>

| Dataset                                                                   |      |
|---------------------------------------------------------------------------|------|
| <a href="https://www.ncbi.nlm.nih.gov/">https://www.ncbi.nlm.nih.gov/</a> | NAME |
| <a href="https://www.ncbi.nlm.nih.gov/">https://www.ncbi.nlm.nih.gov/</a> | LINK |

- 2 Selecciona qué tipo de información registrada se desea acceder (ej. genes, genoma completo, proteínas ..) Nota: Cuando es un fragmento de un gen, dejar "All Databases".



- 3 Colocar la búsqueda del gen o proteína requerido "Search"  
Nota: Para este caso fue un fragmento del DNA ribosomal del nematodo *Toxocara* (ITS-1)

The screenshot shows the NCBI search results for the query 'ITS-1 Toxocara'. The search was performed on the National Library of Medicine website. The results are categorized into four main groups: Literature, Genes, Proteins, and Genomes. The Literature group shows 13 results in PubMed and 108 in PubMed Central. The Genes group shows 0 results in Gene, GEO DataSets, and GEO Profiles. The Proteins group shows 0 results in Conserved Domains, Identical Protein Groups, Protein, Protein Family Models, and Structure. The Genomes group shows 0 results in Assembly / Genome, BioCollections, BioProject, BioSample, Nucleotide, SRA, and Taxonomy. The Clinical group shows 0 results in ClinicalTrials.gov, ClinVar, dbGaP, dbSNP, dbVar, GTR, and MedGen. The PubChem group shows 0 results in BioAssays, Compounds, Pathways, and Substances.

- 4 **Buscar secuencias depositadas (o nucleóticos, o péptidos, o documentos publicados con ese fragmento de DNA en Toxocara ...) Nota: Para este caso "Nucleotide 55"**

The screenshot shows the NCBI search results for the query 'ITS-1%20Toxocara'. The search was performed on the National Library of Medicine website. The results are categorized into four main groups: Literature, Genes, Proteins, and Genomes. The Literature group shows 13 results in PubMed and 108 in PubMed Central. The Genes group shows 0 results in Gene, GEO DataSets, and GEO Profiles. The Proteins group shows 0 results in Conserved Domains, Identical Protein Groups, Protein, Protein Family Models, and Structure. The Genomes group shows 0 results in Assembly / Genome, BioCollections, BioProject, BioSample, Nucleotide, SRA, and Taxonomy. The Clinical group shows 1 result in ClinVar and 0 results in ClinicalTrials.gov, dbGaP, dbSNP, dbVar, GTR, and MedGen. The PubChem group shows 0 results in BioAssays, Compounds, Pathways, and Substances.

- 5 **Identificar la parte amplificada versus lo depositado en NCBI + Secuencia FASTA**



GenBank

**Toxocara cati isolate 20-728 internal transcribed spacer 1, partial sequence**

GenBank: OM822767.1

FASTA Graphics

Go to: ☐

LOCUS OM822767 490 bp DNA linear INV 02-MAR-2022

DEFINITION Toxocara cati isolate 20-728 internal transcribed spacer 1, partial sequence.

ACCESSION OM822767

VERSION OM822767.1

KEYWORDS

SOURCE Toxocara cati (cat roundworm)

ORGANISM Toxocara cati

Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Spirurina; Ascaridomorpha; Ascaridoidea; Toxocaridae; Toxocara.

REFERENCE 1 (bases 1 to 490)

AUTHORS Toce, M., Romano, A.C., Pietragalla, I., Marucci, G. and Palazzo, L.

TITLE Detection of Toxocara cati larvae in a common buzzard (Buteo buteo) and in a red kite (Milvus milvus) in Basilicata Region, Italy

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 490)

AUTHORS Marucci, G.

Send to: Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Related information

Taxonomy

Recent activity

Turn Off Clear

Toxocara cati isolate 20-728 internal transcribed spacer 1, partial sequence

## 6 Dirigirse a "Run BLAST"

GenBank

**Toxocara cati isolate 20-728 internal transcribed spacer 1, partial sequence**

GenBank: OM822767.1

FASTA Graphics

Go to: ☐

LOCUS OM822767 490 bp DNA linear INV 02-MAR-2022

DEFINITION Toxocara cati isolate 20-728 internal transcribed spacer 1, partial sequence.

ACCESSION OM822767

VERSION OM822767.1

KEYWORDS

SOURCE Toxocara cati (cat roundworm)

ORGANISM Toxocara cati

Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Spirurina; Ascaridomorpha; Ascaridoidea; Toxocaridae; Toxocara.

REFERENCE 1 (bases 1 to 490)

AUTHORS Toce, M., Romano, A.C., Pietragalla, I., Marucci, G. and Palazzo, L.

TITLE Detection of Toxocara cati larvae in a common buzzard (Buteo buteo) and in a red kite (Milvus milvus) in Basilicata Region, Italy

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 490)

AUTHORS Marucci, G.

Send to: Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Related information

Taxonomy

Recent activity

Turn Off Clear

Toxocara cati isolate 20-728 internal transcribed spacer 1, partial sequence

ITS-1 Toxocara (56)

o desde la página inicial de BLAST acceder (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) a BLASTn

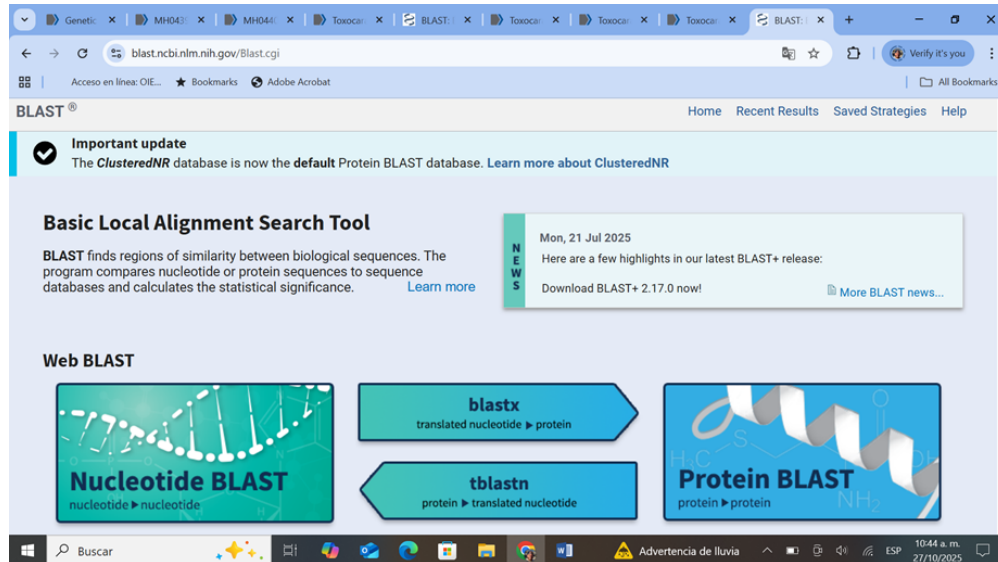
## Dataset

<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

NAME

<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

LINK



7 A. Pegar secuencia FAST + B. Procurar el organismo + C. click "BLAST"

Enter Query Sequence

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

Query subrange [?](#)

From

To

Or, upload file  No file chosen [?](#)

Genetic code

Job Title

☐ Align two or more sequences [?](#)

Choose Search Set

Database  [?](#)

Organism  [Add organism](#)

[+ Algorithm parameters](#)

using a translated nucleotide query

[Feedback](#)

- 8 Esperar cargue de información en página nueva e ir a las secuencias depositadas en la parte inferior de la página

BLAST [®](#) » blastx » results for RID-FZ547MYD014

Home Recent Results Saved Strategies Help

[Important update](#)

The **ClusteredNR** database is now the default Protein BLAST database. [Learn more about ClusteredNR](#)

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

[Your search is limited to records include: Toxocara \(taxid:6264\)](#)

Job Title **Nucleotide Sequence**

RID [FZ547MYD014](#) Search expires on 10-28 23:49 pm [Download All](#)

Program [BLASTX](#) [Citation](#)

Database [ClusteredNR](#) [See details](#)

Query ID [lcl|Query\\_7859978](#)

Description [None](#)

Molecule type [dna](#)

Query Length [490](#)

Other reports [?](#)

Filter Results

Organism [only top 20 will appear](#) [NEW](#)

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity  to

E value  to

Query Coverage  to

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

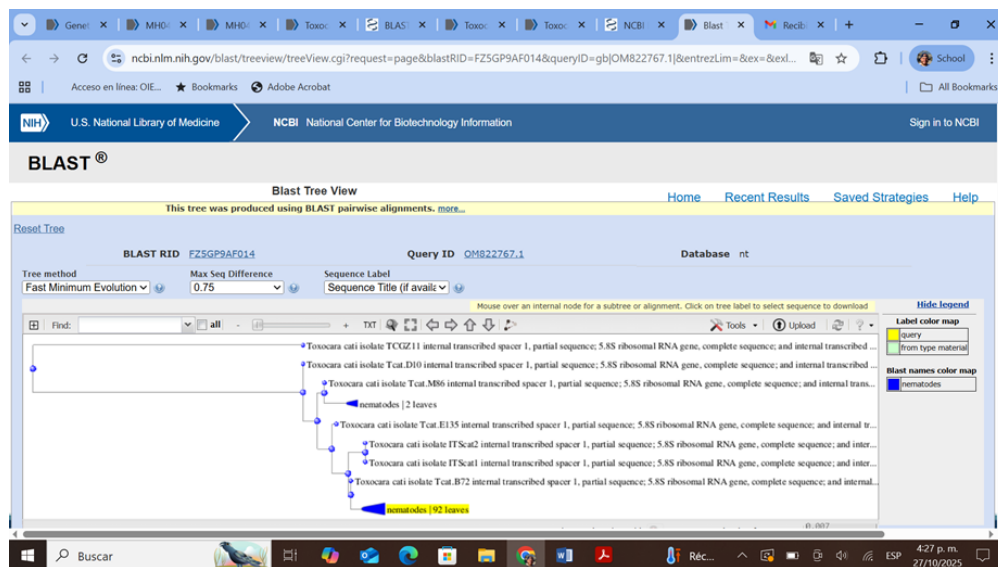
[Feedback](#)

- 9 Seleccionar las secuencias para el estudio genético (ej. a través de árboles filogenéticos)

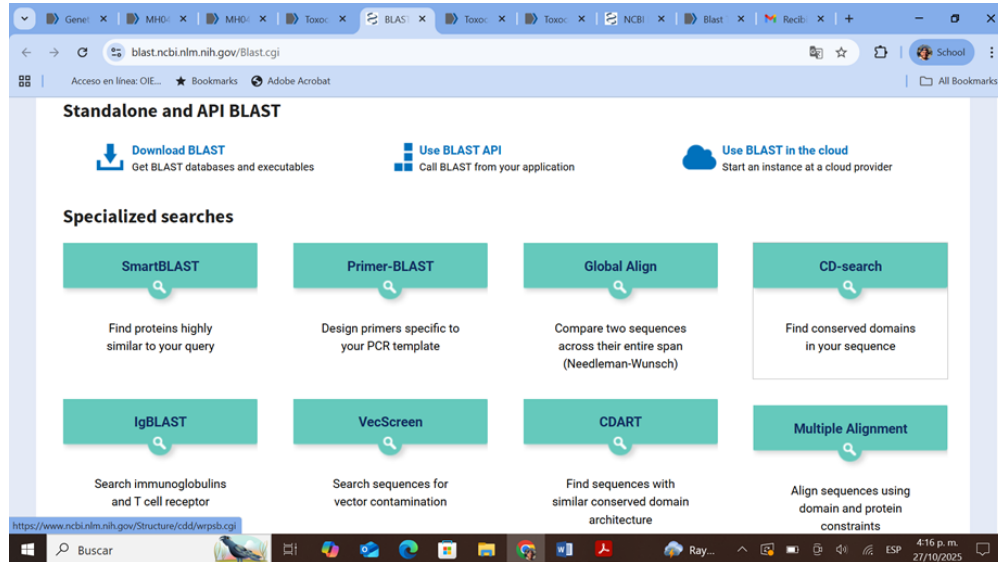
The screenshot shows the NCBI BLAST results page for a query. The table lists sequences with their descriptions, scientific names, and various alignment metrics. The table is titled "Sequences producing significant alignments" and includes columns for Description, Scientific Name, Max Score, Total Score, Query Cover, E value, Per. Ident, Acc. Len, and Accession. The table contains 10 rows of data, all from the Toxocara cati isolate.

| Description                                                                                                      | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession  |
|------------------------------------------------------------------------------------------------------------------|-----------------|-----------|-------------|-------------|---------|------------|----------|------------|
| Toxocara cati isolate No. 768 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer... | Toxocara cati   | 905       | 905         | 100%        | 0.0     | 100.00%    | 554      | PV329859.1 |
| Toxocara cati isolate TCF815 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and...     | Toxocara cati   | 905       | 905         | 100%        | 0.0     | 100.00%    | 1053     | KY003079.1 |
| Toxocara cati isolate 20-728 internal transcribed spacer 1, partial sequence                                     | Toxocara cati   | 905       | 905         | 100%        | 0.0     | 100.00%    | 490      | OM822767.1 |
| Toxocara cati isolate 21-116 internal transcribed spacer 1, partial sequence                                     | Toxocara cati   | 905       | 905         | 100%        | 0.0     | 100.00%    | 490      | OM822766.1 |
| Toxocara cati genes for ITS1, 5.8S rRNA, ITS2, complete and partial sequence, isolate: Tc1                       | Toxocara cati   | 905       | 905         | 100%        | 0.0     | 100.00%    | 979      | AB571303.1 |
| Toxocara cati genes for internal transcribed spacer 1, 5.8S ribosomal RNA, partial sequence, isolate: Tc1        | Toxocara cati   | 902       | 902         | 100%        | 0.0     | 99.80%     | 509      | AB110025.1 |
| Toxocara cati isolate TCJM-19 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and...    | Toxocara cati   | 896       | 896         | 100%        | 0.0     | 99.59%     | 1064     | KY003083.1 |
| Toxocara cati (Australia) first internal transcribed spacer ribosomal DNA                                        | Toxocara cati   | 894       | 894         | 100%        | 0.0     | 99.59%     | 490      | AJ002436.1 |
| Toxocara cati isolate 1 clone 25 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene, co... | Toxocara cati   | 881       | 881         | 100%        | 0.0     | 99.18%     | 978      | KJ777179.1 |
| Toxocara cati isolate Wk072 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and in...   | Toxocara cati   | 876       | 876         | 97%         | 0.0     | 99.79%     | 996      | PP812131.1 |
| Toxocara cati isolate Tc7 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and inter...  | Toxocara cati   | 874       | 874         | 97%         | 0.0     | 99.79%     | 1043     | PV368861.1 |

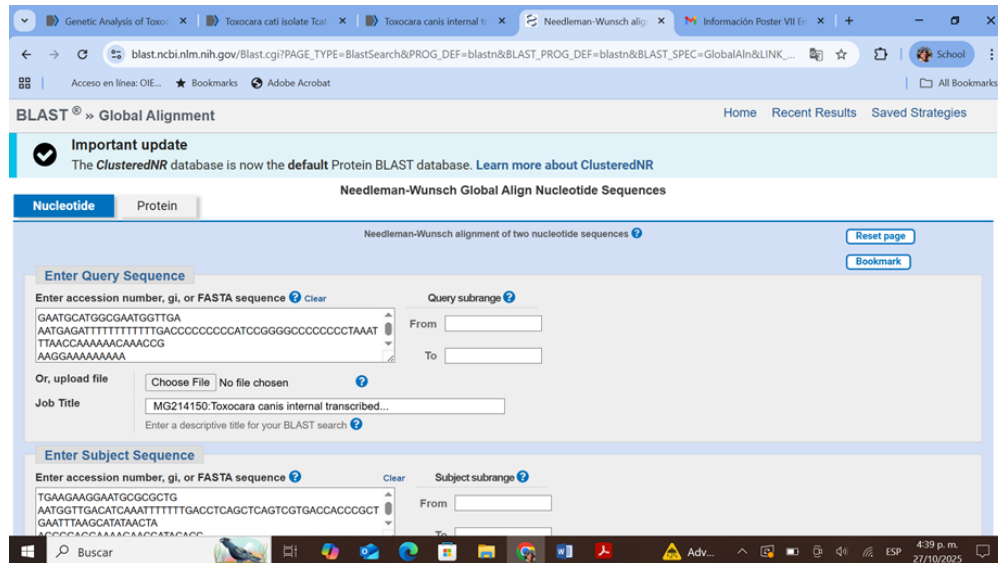
- 10 **BLAST permite generar estudios filogenéticos a través de distanciamiento génico por arboles, esta salida del software permite descargarse en varios tipos de archivos, hasta en PDF.**



- 11 **Otra forma de análisis genético: En la página inicial de BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) acceder al "Global Align"**



- 12 Permite comparar depósitos de secuencias de nucleótidos (Para este caso ITS-2 de *T. canis* versus *T. cati*) + Click "BLAST"



- 13 **Análisis específicos de distanciamiento genético entre dos fragmentos de gen o genes.**



Descriptions Graphic Summary **Alignments** Dot Plot

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

1 sequences selected

[Download](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

Sequence ID: Query\_779403 Length: 374 Number of Matches: 1

Range 1: 1 to 374 [Graphics](#) [Next Match](#) [Previous Match](#)

| NW Score | Identities    | Gaps          | Strand    |
|----------|---------------|---------------|-----------|
| -88      | 290/549 (53%) | 184/549 (33%) | Plus/Plus |

Query 1 MGTACACANSNTRNATRANSACRARTASTNCGCTAGCTGCGATAATAGTGCAG 60  
Sbjct 1 -----GA--AT-GTGC--- 8

Query 61 TTGCAGACACATTGAGCACTAAATTTGGAACGCACATTGCGCATCGGGTTCAATCCCG 120  
Sbjct 9 TT-----AACGCTAA-GCT---TCTGGTGCAATC--- 34

Query 121 TTGSCAGCTCTGGCTGAGGGTCAGTATATTAAAGGATGATGGGCGCGCAATTTATGG 180  
Sbjct 35 ----- 37

Query 181 AATGTGATTAACGCGCAAGGTTGTGGTGATTGCTGGTGAGCTATGCTGGTGGTAAATGGA 240  
Sbjct 73 ----- 77

Se puede graficar el análisis:

NCBI Sequence Viewer 3.51.1 Help Release Notes

lclQuery\_779403 [Link To This View](#) [Feedback](#)

Query\_779403 Find:  Tools Tracks Download ?

Sequence

(U) BLAST Results for: M0214150:Toxocara canis internal transcribed...

Query\_779403: 1..374 (374 nt)

**FOLLOW NCBI**

[Twitter](#) [Facebook](#) [LinkedIn](#) [RSS](#)

Connect with NLM National Library of Medicine 5500 Rockville Pike Web Policies FOIA Help Accessibility

## Protocol references

<https://www.ncbi.nlm.nih.gov/>

<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

## Acknowledgements

A la Dirección General de Investigaciones de la Universidad de los Llanos por proveer los recursos para la estancia corta de investigación del profesor Dumar A. Jaramillo Hernández.