
Basic bioinformatics for gene and protein sequences analysis

**Thewarach Laha, PhD
Department of Parasitology
Faculty of Medicine
KKU**

Contents

- Searching and retrieving DNA and protein sequences from public databases
- Sequence similarity search (Blast search)
- Sequence manipulation (using Bioedit software)

Data retrieving

Application:

Databases

- Pubmed- reference journal
- DNA
- Protein
- Genome

Web site

- NCBI - <http://www.ncbi.nlm.nih.gov/>
- EBI - <http://www.ebi.ac.uk/>
- Sanger centre
<http://www.sanger.ac.uk/>

Practice

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

Retrieving data

- PubMed
- Nucleotide
- Proteins

Search and retrieve data from NCBI

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

The screenshot shows the NCBI website interface. At the top, there is a search bar with a dropdown menu labeled "All Databases" and a red arrow pointing to it with the text "Drop down list". The dropdown menu lists various databases including Assembly, Biocollections, BioProject, BioSample, BioSystems, Books, ClinVar, Conserved Domains, dbGaP, dbVar, Gene, Genome, GEO DataSets, GEO Profiles, GTR, HomoloGene, Identical Protein Groups, MedGen, and MeSH. Below the search bar, there is a navigation menu with "NCBI Home", "Resource List (A-Z)", and "All Resources". The main content area features a "Submit" section with the text "Deposit data or manuscripts into NCBI databases" and a "Download" section with the text "Transfer NCBI data to your computer". On the right side, there is a "Popular Resources" section with a list of resources: PubMed, Bookshelf, PubMed Central, BLAST, Nucleotide, Genome, SNP, Gene, Protein, and PubChem. Red arrows point to PubMed, Nucleotide, Genome, Gene, and Protein in this list. The bottom of the page has a footer with "About the NCBI | Mission | Organization | NCBI News & Blog".

COVID-19 Information

[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)



Nucleotide

The Nucleotide database is a collection of sequences from several sources, including GenBank, RefSeq, TPA and PDB. Genome, gene and transcript sequence data provide the foundation for biomedical research and discovery.

- ### Using Nucleotide
- [Quick Start Guide](#)
 - [FAQ](#)
 - [Help](#)
 - [GenBank FTP](#)
 - [RefSeq FTP](#)

- ### Nucleotide Tools
- [Submit to GenBank](#)
 - [LinkOut](#)
 - [E-Utilities](#)
 - [BLAST](#)
 - [Batch Entrez](#)

- ### Other Resources
- [GenBank Home](#)
 - [RefSeq Home](#)
 - [Gene Home](#)
 - [SRA Home](#)
 - [INSDC](#)

Nucleotide Nucleotide 18s ribosomal RNA AND Opisthorchis Search

Create alert Advanced

Help

COVID-19 Information

Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español

- Species: Animals (35), Customize...
Molecule types: genomic DNA/RNA (34), mRNA (1), Customize...
Source databases: INSDC (GenBank) (35), Customize...
Sequence Type: Nucleotide (34), EST (1)
Sequence length: Custom range...
Release date: Custom range...
Revision date: Custom range...
Clear all
Show additional filters

Summary 20 per page Sort by Default order

Send to: Filters: Manage Filters

See rna RNase I in the Gene database
rna reference sequences Protein (1)

Items: 1 to 20 of 35

<< First < Prev Page 1 of 2 Next > Last >>

- 1. PO226 Opisthorchis viverrini lambda TripEx2 Library Opisthorchis viverrini cDNA 5' similar to Telorchis assula 18S ribosomal RNA gene, partial sequence, GenBank accession number AY222156, mRNA sequence
191 bp linear mRNA
Accession: DR989671.1 GI: 71812280
BioSample Taxonomy
GenBank FASTA Graphics
2. Opisthorchis viverrini from Laos 18S ribosomal RNA gene, partial sequence
1,069 bp linear DNA
Accession: AF408145.1 GI: 15867892
PubMed Taxonomy
GenBank FASTA Graphics
3. Opisthorchis felineus 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, complete sequence; 5.8S ribosomal RNA gene, partial sequence; and internal transcribed spacer 1 and 18S ribosomal RNA gene, complete sequence
78 bp linear DNA
Accession: KR995729.1 GI: 944551871
Taxonomy

Results by taxon

Top Organisms Tree

- Opisthorchis viverrini (17)
Opisthorchis felineus (10)
Opisthorchis pedicellata (2)
Opisthorchis sp. AVK-2013 (2)
Opisthorchis sp. ABG-2016 (1)
All other taxa (3)
More...



Find related data

Database: Select

Find items

Search details

18s ribosomal rna[All Fields] AND ("Opisthorchis"[Organism] OR Opisthorchis[All Fields])

Search

See more...

Recent activity

Nucleotide
[Create alert](#) [Advanced](#) [Help](#)

COVID-19 Information

[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)

- Species
 - Animals (35)
 - Customize ...
- Molecule types
 - genomic DNA/RNA (34)
 - mRNA (1)
 - Customize ...
- Source databases
 - INSDC (GenBank) (35)
 - Customize ...
- Sequence Type
 - Nucleotide (34)
 - EST (1)
- Sequence length
 - Custom range...
- Release date

Summary

Format

- Summary
- GenBank
- GenBank (full)
- FASTA
- FASTA (text)
- ASN.1
- Revision History
- Accession List
- GI List

the Gene database
Protein (1)

5

Opisthorchis viverrini lambda TriplEx2 Library **Opisthorchis viverrini** cDNA 5' similar to
18S ribosomal RNA gene, partial sequence, GenBank accession number

A sequence

191 bp linear mRNA
Accession: DR989671.1 GI: 71812280
[BioSample](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

<< First < Prev Page 1 of 2 Next > Last >>

Results by taxon

Top Organisms [Tree](#)

- Opisthorchis viverrini (17)
- Opisthorchis felineus (10)
- Opisthorchis pedicellata (2)
- Opisthorchis sp. AVK-2013 (2)
- Opisthorchis sp. ABG-2016 (1)
- All other taxa (3)

More...

Find related data

Database:

GenBank
Or
FASTA



Exercise 1

1. Practice- search and retrieve nucleotide and/or protein sequence from NCBI
2. Practice-

Bioinformatics for homology and structure prediction

- Blast homology search

Basic using of NCBI blast

What is Blast?

- The Basic Local Alignment Search Tool (BLAST) is a sequence similarity search tool (probably is the most widely used)
- Finds regions of similarity between 2 sequences (both protein and DNA).
- The program compares DNA or protein sequences to sequence databases and calculates the statistical significance of matches.

What BLAST tells you?

Your query sequence ;

1. What it is related to?
 - Homology
 - Function
2. Is it already in the database?
 - Find the matching sequences in the database
 - Origin of organism
3. Where is it located?
 - In a genome
 - Annotation problem such as frame shifts

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

The image shows a screenshot of the NCBI website homepage. At the top, the browser address bar displays "https://www.ncbi.nlm.nih.gov/". The website header includes the NCBI logo, "National Center for Biotechnology Information", and a search bar with a dropdown menu set to "All Databases". A "Sign in to NCBI" link is visible in the top right corner.

A prominent banner for "COVID-19 Information" is displayed, featuring a warning icon and a close button. Below this, there is a section for "UNITE" (Ending Structural Racism) with a colorful graphic and the text "A new NIH initiative to end structural racism and achieve racial equity in the biomedical research enterprise." A "LEARN MORE" link is provided.

The main content area is divided into three columns:

- Left Column:** A navigation menu with links for "NCBI Home", "Resource List (A-Z)", "All Resources", "Chemicals & Bioassays", "Data & Software", "DNA & RNA", "Domains & Structures", "Genes & Expression", "Genetics & Medicine", "Genomes & Maps", and "Homology".
- Middle Column:** A "Welcome to NCBI" section with the text "The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information." Below this are three main actions: "Submit" (Deposit data or manuscripts into NCBI databases), "Download" (Transfer NCBI data to your computer), and "Learn" (Find help documents, attend a class or watch a tutorial). Each action is accompanied by a corresponding icon.
- Right Column:** A "Popular Resources" section with a list of links: "PubMed", "Bookshelf", "PubMed Central", "BLAST", "Nucleotide", "Genome", "SNP", "Gene", "Protein", and "PubChem". A red arrow points to the "BLAST" link.

■ >AF408145.1 *Opisthorchis viverrini* from Laos 18S ribosomal RNA gene, partial sequence

```
GGTGGATTTATTAGAACAGAACCAACCGGCGGTGACTTCGGTTGCTGTGCGTTGCATTCTGTGATGACTCT
GGATAACTTTACTGATCGCAGTCGGCCTTGTGTCGGCGACGGATCTTTCAAATGTCTGCCCTATCAATTT
TCGATGGTAGGTGACCTGCCTACCATGGTGATAACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGG
GAGCCTGAGAAACGGCTACCACTTCCAAGGAAGGCAGCAGGCGCGCAAATTACCCAATCCCGGCACGGAG
GTAGTGACGAAAAATACGGATACGGGACTCATTAGAGGCTCCGTAATTCGAATGAGTACAATTTAAATCC
TTTAAACGAGGATCAACTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAACTCCAGCTCCAGAAGCGTAT
ATTAAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATCTGGGTCGCATGGCTACATGCCGTTGCTCGTAT
TCCTGGCCTGGTTCACACCGGGACGGGTTTGTGAGTCGGTGTGCTGGTGTGTCAGCCTTTCTGCCGTGTC
TGTTTCGACAGGTGTTAGCATGATTGGTGGGTTTCGTCCTGCTGATCTGTTGACATGCTTCCAGGTGCCCT
TAACCGGGTGTGCGGGGGCGGACGGCACGTTTACTTTGAACAAATTTGAGTGCTCAAAGCAGGCCCGTGTG
CCTGAAAATTCTTGCATGGAATAATGGAATAGGACTTCGGTTCATTTTTGTTGGTTTTTCGGATCCGAAGT
AATGGTTAAAAGGGACAGACGGGGGCATTTGTATGGCGGTGTTAGAGGTGAAATTCTTGGATCACCGCCA
GACAACTACAGCGAAAGCATTTGCCAAGGATGTTTTTCATTGATCTGGAGCGAAAGTCAGAGGTTCTGAAG
ACGATCAGATACCGTCTAGTTCTGACCATAAACGATGCCAACTGACGATCCGTGGTGGCGCGATTATTG
GCCCCGCGGGCAGCCCCCGGAAACCTTTAAGTCTTTGGGCTCCGGGGGAAGTATGGTTGCAAAGCTGAA
ACTTAAATGAATTGACGGA
```

ATGGTCTCGCTAAGTTGTGGATATCGTTGTCTTCAAATTATACTTGTTCATCCTTAATACTTTTGTTCATCG
CATGTGGTGTGGCCTCATAGTTGTCGGGAGTTTGGCCGAAGTTTCTTTGAAGACTTTCGGAGAATCTAA
CGAAACCAGCATCCAAATTATCGTTATTTTCATCATTGTCTCGGATGTCTCACATTTTAAATAGGATTT
TTGGGGTTCTGTGGAGCATGCCTGAAAAATGTCTGTATGCTCATTACCTATGCTATACTGCTTGGAGTGA
CGGCAGTTGCACAAGTCGTATGTGGAATCATTGGCCTTGTTCTCCGCGATAAGATCCCCGGATTGGTAAA
TCACAATCTGGAGGTGCTGTATACAGAGTATAGTGCCAACAAAGATGTTCAAAGTTGATTAACGTGATC
CAGAGTGAGCTGAAATGCTGTGGAGCGACTGGAACGTGGGCGAACCCGGGGTCTGAACCAGAATCTTGCC
GCAGCCCAGAAGGTGTCGTTTACAAGGATGGCTGTGTTCCGAAAGTCGAAGAGTTCATCCAGGAAAATAT
GGTTGCAATCGGCGTGTGCGTCTTCATTTTTGCACTAATTCAACTCATCTGTATGACGTTTGCCATCTGT
GTTGTGCAAGCTTTACGAAAAGGAGAAGGCGAGACGGTATAG

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](#)

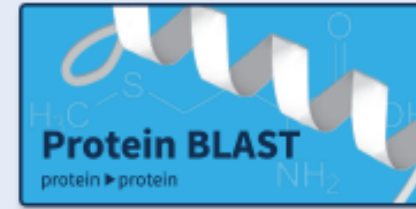
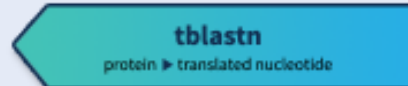
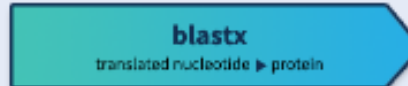
ElasticBLAST is here!

ElasticBLAST is a new cloud based tool to run your BLAST searches faster and make you more effective.

Mon, 07 Feb 2022 12:00:00 EST

[More BLAST news...](#)

Web BLAST



BLAST Genomes

[Human](#)

[Mouse](#)

[Rat](#)

[Microbes](#)

Standalone and API BLAST



Download BLAST

Get BLAST databases and executables



Use BLAST API

Call BLAST from your application



Use BLAST in the cloud

Start an instance at a cloud provider

Specialized searches

SmartBLAST



Find proteins highly similar to your query

Primer-BLAST



Design primers specific to your PCR template

Global Align



Compare two sequences across their entire span (Needleman-Wunsch)

CD-search



Find conserved domains in your sequence

IgBLAST



Search immunoglobulins and T cell receptor

VecScreen



Search sequences for vector contamination

CDART



Find sequences with similar conserved domain

Multiple Alignment



Align sequences using domain and protein

Functions of BLAST Search Pages

1. Nucleotide blast- **blastn suite**

- megablast: for sequence identification, intra-species comparison
- discontinuous megablast: for cross-species comparison, searching with coding sequences
- blastn: for searching with shorter queries, cross-species comparison

Protein blast

- **blastp**: general sequence identification and similarity searches
- **blastx**: for identifying potential protein products encoded by a nucleotide query
- **tblastn**: for identifying database sequences encoding proteins similar to the query
- **tblastx**: for identifying nucleotide sequences similar to the query based on their coding potential

Table 1. Key features of the BLAST search pages in the “Basic BLAST” category

Search page	Query & database combination	Alignment type	Programs & functions (default program in bold)
nucleotide blast	nucleotide vs nucleotide	nucleotide vs nucleotide	megablast : for sequence identification, intra-species comparison <u>discontiguous megablast</u> : for cross-species comparison, searching with coding sequences <u>blastn</u> : for searching with shorter queries, cross-species comparison
protein blast	Protein vs protein	protein vs protein	blastp : general sequence identification and similarity searches <u>DELTA-BLAST</u> [2] : protein similarity search with higher sensitivity than blastp <u>PSI-BLAST</u> : iterative search for position-specific score matrix (PSSM) construction or identification of distant relatives for a protein family <u>PHI-BLAST</u> : protein alignment with input pattern as anchor/constraint
blastx	nucleotide (translated) vs protein	protein vs protein	<u>blastx</u> : for identifying potential protein products encoded by a nucleotide query
tblastn	protein vs nucleotide (translated)	protein vs protein	<u>tblastn</u> : for identifying database sequences encoding proteins similar to the query
tblastx	nucleotide (translated) vs nucleotide (translated)	protein vs protein	<u>tblastx</u> : for identifying nucleotide sequences similar to the query based on their coding potential

COVID-19 Information

[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)

BLAST® » **blastn suite**

Home Recent Results Saved Strategies Help



Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

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

Reset page Bookmark

Enter Query Sequence

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

Query subrange [?](#)

From

To

Or, upload file

Choose File No file chosen [?](#)

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) [?](#)

 **COVID-19 Information** 


[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)

Standard Nucleotide BLAST


- blastn**
- blastp
- blastx
- tblastn
- tblastx

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

Enter Query Sequence


Enter accession number(s), gi(s), or FASTA sequence(s) 

```
ATGGTCTCGCTAAGTTGTGGATATCGTTGTCTTCAAATTATACTTGTCAATCC
TTAATACTTTTGTCAATCC
CATGTGGTGTGGCCCTCATAGTTGTCCGGAGTTTGGCCGAAGTTTCTTTG
AAGACTTTCGGAGAATCTAA
```


Query subrange 


From

To

Or, upload file No file chosen 


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



AAGACTTTTCGGAGAAATCTAA

To

Or, upload file No file chosen

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 Retrovirus

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

Result from BLAST search

blast.ncbi.nlm.nih.gov/Blast.cgi

Query Length 1069

Other reports [Distance tree of results](#) [MSA viewer](#) ?

Filter Reset

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments

Download **New** Select columns Show 100 ?

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

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Opisthorchis viverrini from Laos 18S ribosomal RNA gene, partial sequence	Opisthorchis vive...	1975	1975	100%	0.0	100.00%	1069	AF408145.1
<input checked="" type="checkbox"/>	Clonorchis sinensis isolate Korea 18S ribosomal RNA gene, partial sequence	Clonorchis sinensis	1969	1969	100%	0.0	99.91%	1070	AF217093.1
<input checked="" type="checkbox"/>	Metorchis orientalis isolate MOE 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S...	Metorchis orientalis	1960	1960	100%	0.0	99.72%	9162	MK482055.1
<input checked="" type="checkbox"/>	Metorchis orientalis isolate MOD 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S...	Metorchis orientalis	1960	1960	100%	0.0	99.72%	9355	MK482054.1
<input checked="" type="checkbox"/>	Metorchis orientalis isolate MOC 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S...	Metorchis orientalis	1960	1960	100%	0.0	99.72%	8500	MK482053.1
<input checked="" type="checkbox"/>	Metorchis orientalis isolate MOB 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S...	Metorchis orientalis	1960	1960	100%	0.0	99.72%	9162	MK482052.1
<input checked="" type="checkbox"/>	Metorchis orientalis isolate MOA 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S...	Metorchis orientalis	1960	1960	100%	0.0	99.72%	7881	MK482051.1
<input checked="" type="checkbox"/>	Clonorchis sinensis isolate CSE 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S...	Clonorchis sinensis	1960	1960	100%	0.0	99.72%	8049	MK450527.1
<input checked="" type="checkbox"/>	Clonorchis sinensis isolate CSC 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S...	Clonorchis sinensis	1960	1960	100%	0.0	99.72%	8391	MK450525.1
<input checked="" type="checkbox"/>	Clonorchis sinensis isolate CSB 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S...	Clonorchis sinensis	1960	1960	100%	0.0	99.72%	8391	MK450524.1
<input checked="" type="checkbox"/>	Clonorchis sinensis isolate CSA 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S...	Clonorchis sinensis	1960	1960	100%	0.0	99.72%	8049	
<input checked="" type="checkbox"/>	Opisthorchis felinus strain Tula small subunit ribosomal RNA gene, partial sequence	Opisthorchis feli	1960	1960	100%	0.0	99.72%	1960	

[Feedback](#)

Practice

- Run BLAST search with your interested sequences.

Basic tools

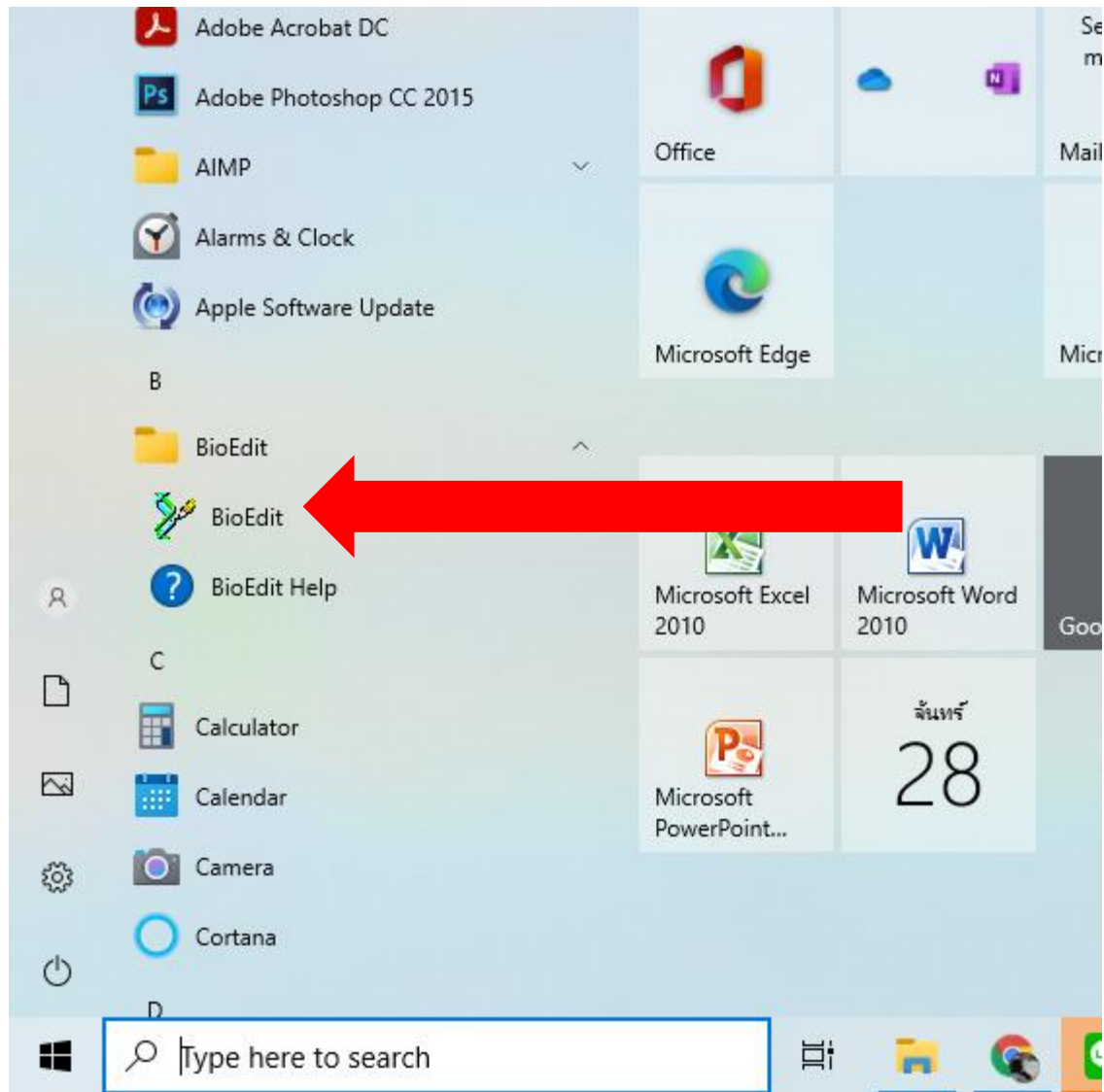
BioEdit Software

- **Sequence editing**
 - View and edit DNA /protein sequence file
- **Sequence manipulation**
 - DNA or nucleic acid sequence
 - Base composition- %CG
 - Complement or reverse complement
 - Restriction enzyme map
 - Translation to protein (amino acid sequence)
 - Protein
 - Amino acid composition- prediction of MW
- **Sequence alignment**

Practice

1. Install BioEdit software

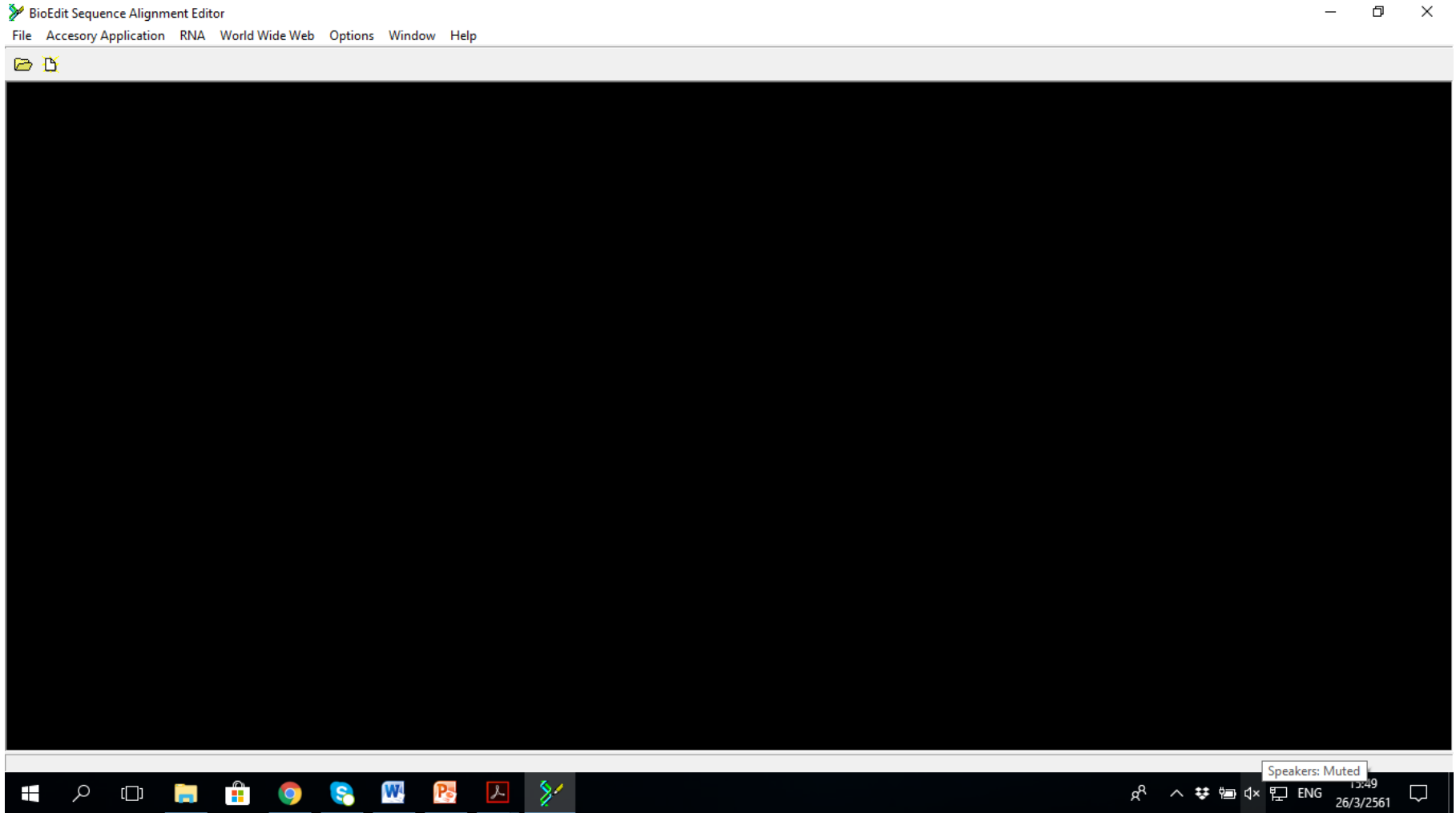
- Google search- BioEdit software download.
- Download BioEdit (compatible version for your OS and computer).
- Unzip and install program.



BioEdit Software

- **Sequence editing**
 - **View and edit DNA /protein sequence file**

Open program



- New Alignment Ctrl+N
- Open... Ctrl+O
- Open motif descriptor
- New from Clipboard
- Retrieve sequences from GenBank or GenPept
- Encrypt file with a .jpg image and encode into proteins
- New Text
- Open As Text
- Run script

- Batch ABI to SCF Trace File Conversion
- Batch Export of Raw Sequence Trace Data
- Batch Export of (artificially) Reverse-Complemented Raw Sequence Trace Data

- D:\OV_EXOSOME_Project\TSP2 project\TSP2_manuscript\input sequences.gb
- D:\OV_EXOSOME_Project\TSP2 project\TSP2_manuscript\alignment.gb
- Untitled1
- C:\BioEdit\Temp\~out.tmp

- Exit

Sequence editing



BioEdit Software

- **Sequence manipulation**
 - **DNA or nucleic acid sequence**
 - **Base composition- %CG**
 - **Complement or reverse complement**
 - **Restriction enzyme map**
 - **Translation to protein (amino acid sequence)**
 - **Protein**
 - **Amino acid composition- prediction of MW**

- New Sequence
- Edit Sequence
- Edit all selected
- Select Positions
- Open at cursor position
- Extract Positions
- Go to Pubmed references
- Overwrite/Retrieve sequences by gi number in title by HTTP to GenBank using gi|###|
- Overwrite/Retrieve Genbank data by gi number in title by HTTP to GenBank (ignore sequence) using gi|###|
- Overwrite/Retrieve sequences by gi number in title by HTTP to GenBank using gb|<accession>
- Overwrite/Retrieve Genbank data by gi number in title by HTTP to GenBank (ignore sequence) using gb|<accession>
- True positions from alignment positions
- Phylogeny / Taxonomy >
- Filter out sequences containing certain characters >
- Rename >
- Sort >
- PCR Primers / oligos >
- Pairwise alignment >
- Similarity Matrix (for pairwise alignments and shading) >
- Features >
- Sequence groups (or families) >
- Edit Mode >
- Mask >
- Toggle Color
- Gaps >
- Manipulations >
- Nucleic Acid >
- Protein >
- Translate or Reverse-Translate (permanent)
- Translate in selected frame (permanent)
- Toggle Translation

Ctrl+G

Sequence manipulation (DNA)

The image shows a screenshot of the BioEdit Sequence Alignment Editor interface. The main window displays a DNA sequence with a selection box around a portion of it. A menu is open, showing the 'Sequence' menu and its 'Nucleic Acid' submenu. A red arrow points from the 'Functions' label in the sequence editor to the 'Nucleic Acid' submenu.

Sequence Editor Menu:

- New Sequence
- Edit Sequence
- Edit all selected
- Select Positions
- Open at cursor position
- Extract Positions
- Go to Pubmed references
- Overwrite/Retrieve sequences by gi number in title by HTTP to GenBank using gi|###|
- Overwrite/Retrieve Genbank data by gi number in title by HTTP to GenBank (ignore sequence) using gi|###|
- Overwrite/Retrieve sequences by gi number in title by HTTP to GenBank using gb|<accession>
- Overwrite/Retrieve Genbank data by gi number in title by HTTP to GenBank (ignore sequence) using gb|<accession>|
- True positions from alignment positions
- Phylogeny / Taxonomy
- Filter out sequences containing certain characters
- Rename
- Sort
- PCR Primers / oligos
- Pairwise alignment
- Similarity Matrix (for pairwise alignments and shading)
- Features
- Sequence groups (or families)
- Edit Mode
- Mask
- Toggle Color
- Gaps
- Manipulations
- Nucleic Acid**
- Protein
- Translate or Reverse-Translate (permanent)
- Translate in selected frame (permanent)
- Toggle Translation

Nucleic Acid Submenu:

- Nucleotide Composition
- Base composition and mass export (monoisotopic)
- Base composition and mass export with average masses
- Complement
- Reverse Complement Shift+Ctrl+R
- DNA->RNA
- RNA->DNA
- Translate
- Find next ORF
- Find ORFs from a list of positions
- Create Plasmid from Sequence
- Gap beginning to minimize stop codons in reading frame 1 codon search
- Restriction Map
- Sorted Six-Frame Translation
- Unsorted Six-Frame Translation

Sequence Editor Content:

Opisthorchis atgtc
Sm_LAP ATGAC

Functions

90 100 110 120 130

tgtaatcagttgaaactccagcgatcacgactgtcttgcctggttaccgatg
CATTACCTAATGCTTACAGCTGGCTTTGAATGCGTTAAAGGAATTTCTCG

Sequence manipulation (protein)

The image shows a screenshot of the BioEdit Sequence Alignment Editor. The 'Sequence' menu is open, and the 'Protein' option is selected. A sub-menu is displayed, listing various protein analysis functions. A red arrow points from the 'Function' label to the sub-menu.

Function

- Amino Acid Composition
- Helical Wheel Diagram
- Hydrophobic Moment matrix with Eisenberg consensus scale
- Kyte & Doolittle Mean Hydrophobicity Profile
- Eisenberg Scale Mean Hydrophobicity Profile
- Cornette Scale Mean Hydrophobicity Profile
- Parker HPLC Scale Mean Hydrophilicity Profile
- Boyko Scale Mean Hydrophilicity Profile
- Hopp & Woods Scale Mean Hydrophilicity
- Eisenberg Hydrophobic Moment Profile
- Mean Eisenberg Hydrophobic Moment Profile

The background shows the BioEdit interface with a sequence editor window. The sequence being edited is:

```
Opisthorchis atgtc
Sm_LAP ATGAC
```

The sequence editor window also shows a scale from 90 to 130 and a sequence: `tgtaatcagttgaactccagcgatcagcactgtcttgcctggttaccgatg` and `CATTACCTAATGCTTTACAGCTGGCTTTGAAATGCGTTAAAGGAATTTCTG`.

Sequences manipulation

DNA sequence

- **Translation:** translate DNA to protein (amino acid) sequence
 - BioEdit-sequence-nucleic acid-translate
 - Web
<http://bio.lundberg.gu.se/edu/translate.html>

- Reverse complement
- Complement
- Contig assembly

File Edit Sequence

Untitled

Courier New

Mode: Select / Slide

I D I D

ตัวเรียงลำดับภาพนิ่ง *Bold

Type he

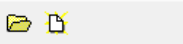
- New Sequence
- Edit Sequence
- Edit all selected
- Select Positions
- Open at cursor position
- Extract Positions
- Go to Pubmed references
- Overwrite/Retrieve sequences by gi number in title by HTTP to GenBank using gi|###|
- Overwrite/Retrieve Genbank data by gi number in title by HTTP to GenBank (ignore sequence) using gi|###|
- Overwrite/Retrieve sequences by gi number in title by HTTP to GenBank using gb|<accession>
- Overwrite/Retrieve Genbank data by gi number in title by HTTP to GenBank (ignore sequence) using gb|<accession>|
- True positions from alignment positions
- Phylogeny / Taxonomy >
- Filter out sequences containing certain characters >
- Rename >
- Sort >
- PCR Primers / oligos >
- Pairwise alignment >
- Similarity Matrix (for pairwise alignments and shading) >
- Features >
- Sequence groups (or families) >
- Edit Mode >
- Mask >
- Toggle Color
- Gaps >
- Manipulations >
- Nucleic Acid >**
- Protein >
- Translate or Reverse-Translate (permanent)
- Translate in selected frame (permanent)
- Toggle Translation

- Nucleotide Composition
- Base composition and mass export (monoisotopic)
- Base composition and mass export with average masses
- Complement
- Reverse Complement Shift+Ctrl+R
- DNA->RNA
- RNA->DNA
- Translate >
- Find next ORF
- Find ORFs from a list of positions
- Create Plasmid from Sequence
- Gap beginning to minimize stop codons in reading frame 1 codon search
- Restriction Map
- Sorted Six-Frame Translation
- Unsorted Six-Frame Translation

Ctrl+G

Sequence manipulation (cont.)

- Protein sequence



Untitled

Courier New

Mode: Select / Slide



- New Sequence
- Edit Sequence
- Edit all selected
- Select Positions
- Open at cursor position
- Extract Positions
- Go to Pubmed references
- Overwrite/Retrieve sequences by gi number in title by HTTP to GenBank using gi|###|
- Overwrite/Retrieve Genbank data by gi number in title by HTTP to GenBank (ignore sequence) using gi|###|
- Overwrite/Retrieve sequences by gi number in title by HTTP to GenBank using gb|<accession>
- Overwrite/Retrieve Genbank data by gi number in title by HTTP to GenBank (ignore sequence) using gb|<accession>
- True positions from alignment positions
- Phylogeny / Taxonomy >
- Filter out sequences containing certain characters >
- Rename >
- Sort >
- PCR Primers / oligos >
- Pairwise alignment >
- Similarity Matrix (for pairwise alignments and shading) >
- Features >
- Sequence groups (or families) >
- Edit Mode >
- Mask >
- Toggle Color
- Gaps >
- Manipulations >
- Nucleic Acid >
- Protein >**
- Translate or Reverse-Translate (permanent)
- Translate in selected frame (permanent)
- Toggle Translation

- Amino Acid Composition
- Helical Wheel Diagram
- Hydrophobic Moment matrix with Eisenberg consensus scale
- Kyte & Doolittle Mean Hydrophobicity Profile
- Eisenberg Scale Mean Hydrophobicity Profile
- Cornette Scale Mean Hydrophobicity Profile
- Parker HPLC Scale Mean Hydrophilicity Profile
- Boyko Scale Mean Hydrophilicity Profile
- Hopp & Woods Scale Mean Hydrophilicity
- Eisenberg Hydrophobic Moment Profile
- Mean Eisenberg Hydrophobic Moment Profile

Multiple sequence alignment

1. Retrieve DNA or protein sequence from GenBank (more than 3 sequences).
2. Sequence should be high similarity (same gene or protein)
3. Save in BioEdit sequence file
4. Align-
 - Accessory application-ClustalW multiple alignment

Practice

1. Retrieve DNA sequence from GenBank database (Select interested sequences)
2. Import new sequence into BioEdit
3. Practice
 - Manipulate sequence
 - Multiple sequences alignment



LiEdit Sequence Alignment Editor

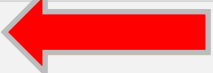
File Accessory Application RNA World Wide Web Options Window Help

- New Alignment Ctrl+N
- Open... Ctrl+O
- Open motif descriptor
- New from Clipboard
- Retrieve sequences from GenBank or GenPept
- Encrypt file with a .jpg image and encode into proteins
- New Text
- Open As Text
- Run script

- Batch ABI to SCF Trace File Conversion
- Batch Export of Raw Sequence Trace Data
- Batch Export of (artificially) Reverse-Complemented Raw Sequence Trace Data

- D:\Luyen\Thesis_Luyen\OV_LAP\lx187340.gb
- D:\OV_EXOSOME_Project\TSP2 project\TSP2_manuscript\input sequences.gb
- D:\Golf\PhD project\Cat B project\Manuscript\CBTree\alignment for CB tree_ThL.phy
- D:\Sirikanda PhD project\Alignment for Lin poster.phy

- Exit





Untitled

0 total sequences

Courier New 11 B

Mode: Select / Slide Selection: 0 Position: Sequence Mask: None Numbering Mask: None Start ruler at: 1

⏏ I D I U G-D + [Colorful icons] MI [Colorful icons] Scroll speed slow fast

10 20 30 40 50 60 70 80

[Empty text area]



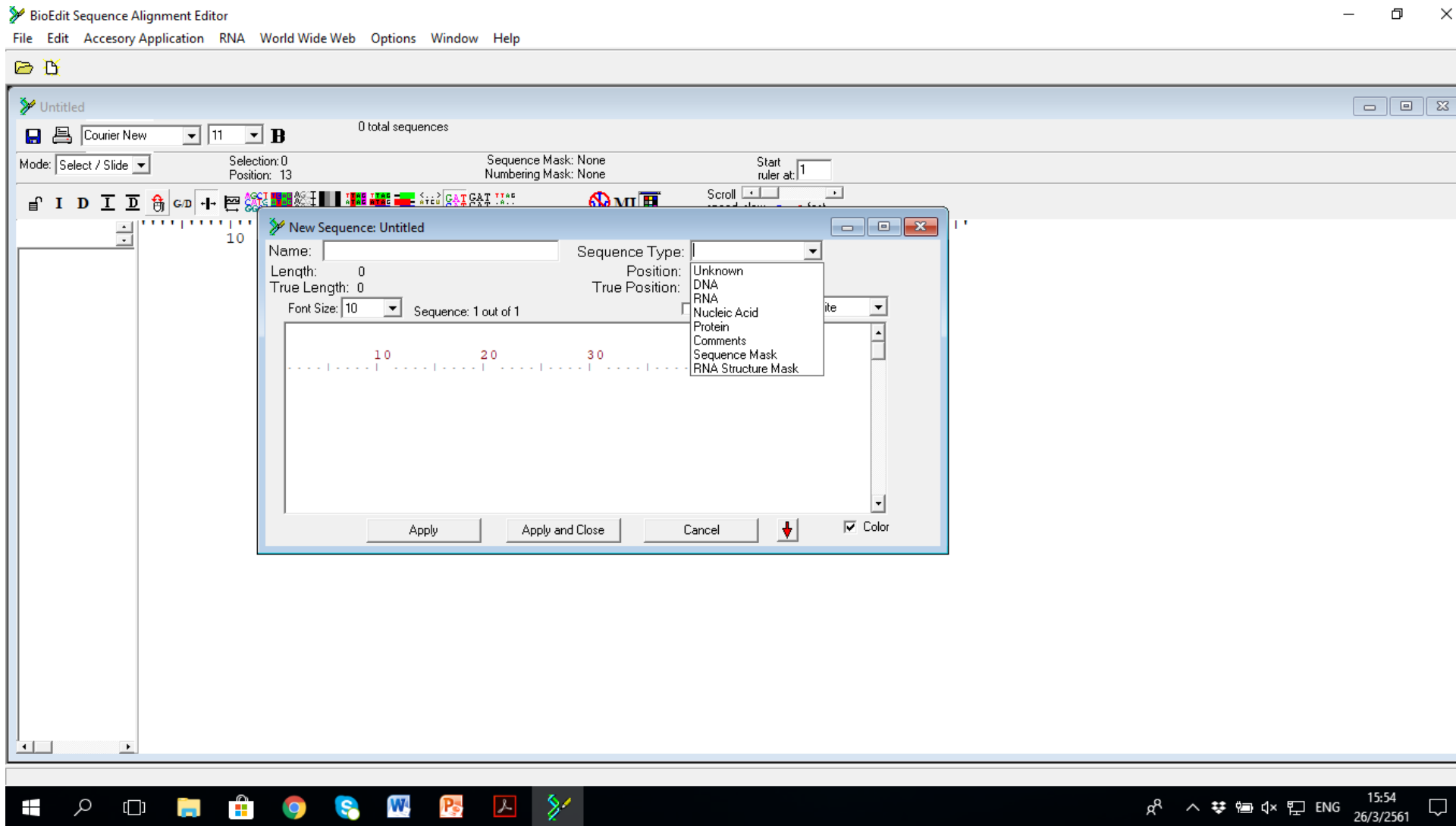
New sequence

The image shows a screenshot of the BioEdit Sequence Alignment Editor software. The 'Sequence' menu is open, displaying a list of options. The menu items are:

- New Sequence
- Edit Sequence
- Edit all selected
- Select Positions
- Open at cursor position
- Extract Positions
- Go to Pubmed references
- Overwrite/Retrieve sequences by gi number in title by HTTP to GenBank using gi|###|
- Overwrite/Retrieve Genbank data by gi number in title by HTTP to GenBank (ignore sequence) using gi|###|
- Overwrite/Retrieve sequences by gi number in title by HTTP to GenBank using gb|<accession>
- Overwrite/Retrieve Genbank data by gi number in title by HTTP to GenBank (ignore sequence) using gb|<accession>|
- True positions from alignment positions
- Phylogeny / Taxonomy >
- Filter out sequences containing certain characters >
- Rename >
- Sort >
- PCR Primers / oligos >
- Pairwise alignment >
- Similarity Matrix (for pairwise alignments and shading) >
- Features >
- Sequence groups (or families) >
- Edit Mode >
- Mask >
- Toggle Color >
- Gaps >
- Manipulations >
- Nucleic Acid >
- Protein >
- Translate or Reverse-Translate (permanent)
- Translate in selected frame (permanent)
- Toggle Translation

The software interface includes a menu bar with 'File', 'Edit', and 'Sequence'. The 'Sequence' menu is currently active. The main window shows a sequence editor with a toolbar and a text area. The Windows taskbar at the bottom displays the time as 15:53 and the date as 26/3/2561.

Paste sequence in box, name sequence and indicate sequence type.



The screenshot shows the BioEdit Sequence Alignment Editor interface. A window titled "Edit Opisthorchis viverrini leucine aminopeptidase (LAP-1) mRNA; D:\Luyen\Thesis_Luyen\O..." is open, displaying a DNA sequence. The sequence is shown in a grid format with positions 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, and 150 marked. The sequence text is:
atgtctgtga gccgtagcgt cactggagat tcgggggtga acgcgcaaga
acagggcatt tgtgccataa cagtacccat caatgtgtgt aatcagttga
actccagcga tcacgactgt cttgtcctgg ttaccgatga cgtatccatc
A red arrow points from a red box labeled "Paste sequence here" to the sequence text. Another red arrow points from the bottom of the window to the "Apply and Close" button.

Paste sequence here

Apply

Apply and Close

Cancel

Color

Add more sequences

The screenshot displays the BioEdit Sequence Alignment Editor interface. The title bar reads "BioEdit Sequence Alignment Editor - [D:\Luyen\Thesis_Luyen\OV_LAP\lx187340.gb]". The menu bar includes "File", "Edit", "Sequence", "Alignment", "View", "Accessory Application", "RNA", "World Wide Web", "Options", "Window", and "Help". The toolbar shows various editing and alignment tools. The main window displays two sequences: "Opisthorchis" and "Sm_LAP". The "Opisthorchis" sequence is highlighted in black. The "Sm_LAP" sequence is shown below it, with a ruler above indicating positions from 10 to 130. The sequence text is: `atgtctgtgagccgtagcgtcaactggagattcgggggtgaacgcgcaagaaacagggcatttggtgccataacagtacccatcaatgtgtgtaatcagttgaactccagcgatcacgaactgtcttgcctggttaccgatg
ATGAGTCTCGTAACCTCCTGCCCTCCAGTCTTTAATCTGGTGGATTCAGCATAATGATGCAGTTTTGCTACTAGACGACGATGTAGAACATTTACCTAATGCTTTACAGCTGGCTTTGAATGCGTTAAAGGAATTTTCTG`. The status bar at the bottom shows the system tray with the time 16:00 and date 26/3/2561.

Sequence alignment

- Pairwise alignment- เป็นการ align sequence 2 sequence เพื่อแสดง identities และ similarities

Add sequences

BioEdit Sequence Alignment Editor - [D:\OV_EXOSOME_Project\TSP2 project\TSP2_manuscript\input sequences.gb]

File Edit Sequence Alignment View Accessory Application RNA World Wide Web Options Window Help

8 total sequences

Mode: Select / Slide Selection: 0 Position: Sequence Mask: None Numbering Mask: None Start ruler at: 1

Scroll speed: slow fast

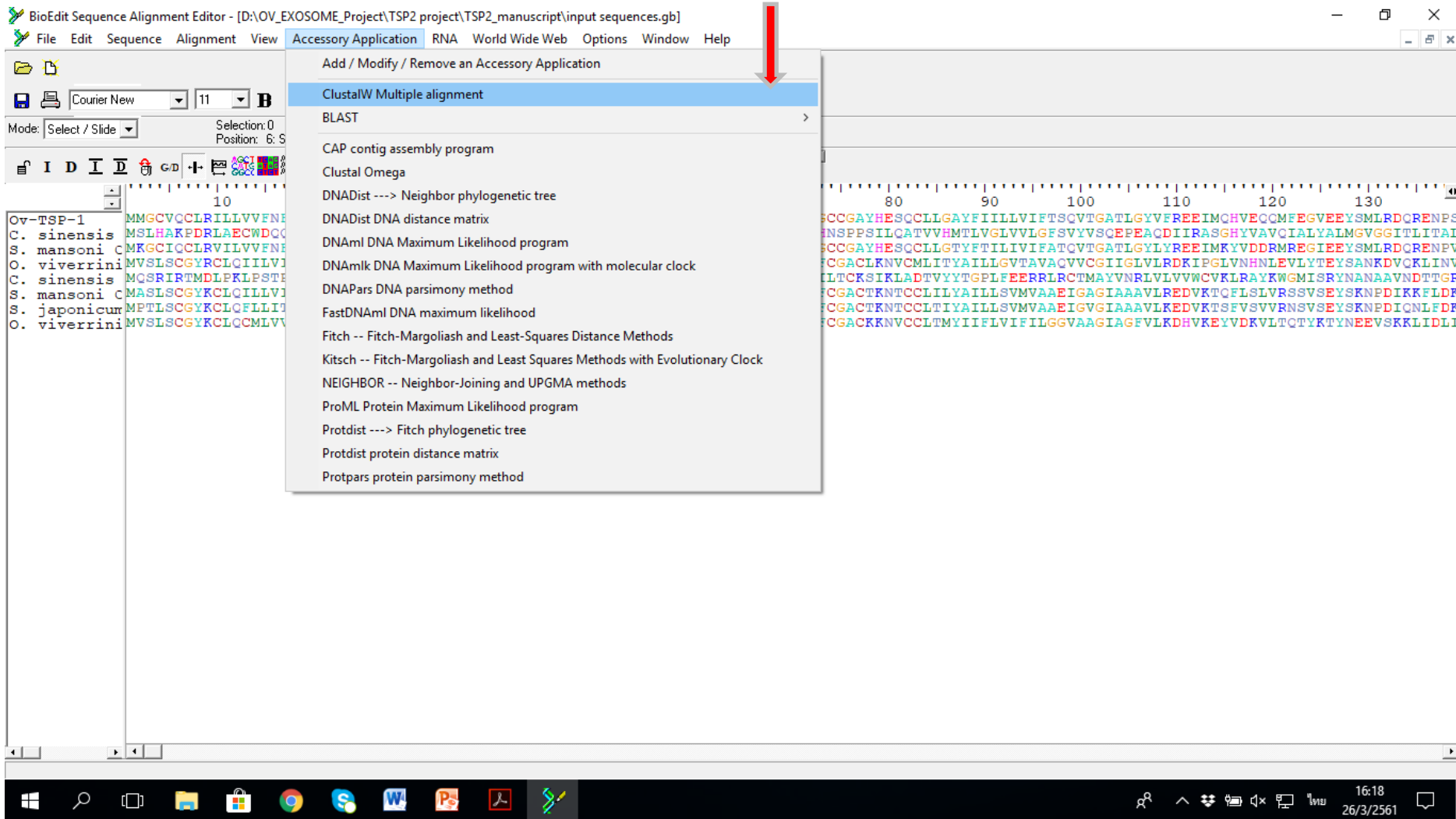
10 20 30 40 50 60 70 80 90 100 110 120 130

```
Ov-TSP-1 MMGCVQCLRILLVVFNFLVVLVGLVVLGFSVYVVSQEPPEAQDIIRASGHYVAVQIALYALMGVGGITLITALFGCCGAYHESQCLLGAYFIILLVIFTSQVTPGATIGYVFREEIMCHVEEQCMFEGVEEYSMLRDQRENPS
C. sinensis MSLHAKPDRLEACWDQQLWDTEMLITVGRQEAAGSCCVDFLAAIRSADLKEIQCSLNRVRWAGQAVLEAESLHNSPPSILQATVVHMTLVGLVVLGFSVYVVSQEPPEAQDIIRASGHYVAVQIALYALMGVGGITLITAL
S. mansoni MKGCIQCLRVLVILVFNFLVVLGFSVYVYIYLEPEAQDIILASGQCHAVQIMLYALMGIGGITLVIALFGCCGAYHESQCLLGTYFTILIVIFATQVTGATLGYLRYEEIMKYVDDRREGIEEYSMLRDQRENPV
O. viverrini MVSLSCGYRCLQIILVILNTEFVIACGVGLIVVGSLEEVSLKTFGESNETSIQIIVIFIICLGLTFLIGFLGFCGACLKNVCMILTYAAILLGVTAQVAVCGIIGLVLRDKIPGLVNHLEVLVTEYSANKDVKQLINLV
C. sinensis MQSRIRTMDLKPLPSTPTFGEASAPVTTVVCTLWKHGQTVILRRQVFTTISVDRKQKQSSWAKPSKQGGTSPILTCRSIKLADTVVYTGPLFEERRLRCTMAYVNRVLVVLVWCVKLRAYKGMISRYNANAANDTGR
S. mansoni MASLSGCKLQILLVIFNIIIVFACGIALIVIGSLSQVAINNYSSGIDSNIKGLVIFIIIVLGCFLFLLGFLGFCGACTKNTCCLILYAILLSVMVAEEIGAGIAAAVLRDVKVQFLSLVRSVSEYSKNPDIKKFLDK
S. japonicum MPTLSCGYKCLQFLITFNIIIVFACGIALIVIGSLSQVAVNNYSTGIDSNIKGLVIFIIIVLGCFLFLLGFLGFCGACTKNTCCLTIYAILLSVMVAEEIGVIAAAVLRDVKVTSFVSVVRNSVSEYSKNPDIQNLFDK
O. viverrini MVSLSCGYKCLQCMVLFVFNVVVICCGIALIVVGSIAQVQLKTYLSSEDAQLMAFVIFIIIFGCFELTVVGSFGFCGACKKNVCCLTMYIIFLVIFILGGVAAGIAGFVLRDVKVYVVDKVLVTQTYKTYNEEVSKKLIDL
```

Windows taskbar: 16:17 26/3/2561

1. Accessory application

2. ClustalW Multiple alignment



The screenshot shows the BioEdit Sequence Alignment Editor interface. The title bar reads "BioEdit Sequence Alignment Editor - [D:\OV_EXOSOME_Project\TSP2 project\TSP2_manuscript\input sequences.gb]". The menu bar includes "File", "Edit", "Sequence", "Alignment", "View", "Accessory Application", "RNA", "World Wide Web", "Options", "Window", and "Help". The "Accessory Application" menu is open, displaying a list of options. A red arrow points to the "ClustalW Multiple alignment" option, which is highlighted in blue. Other options in the menu include "BLAST", "CAP contig assembly program", "Clustal Omega", "DNADist ----> Neighbor phylogenetic tree", "DNADist DNA distance matrix", "DNAMl DNA Maximum Likelihood program", "DNAMl DNA Maximum Likelihood program with molecular clock", "DNAPars DNA parsimony method", "FastDNAMl DNA maximum likelihood", "Fitch -- Fitch-Margoliash and Least-Squares Distance Methods", "Kitsch -- Fitch-Margoliash and Least Squares Methods with Evolutionary Clock", "NEIGHBOR -- Neighbor-Joining and UPGMA methods", "ProML Protein Maximum Likelihood program", "Protodist ----> Fitch phylogenetic tree", "Protodist protein distance matrix", and "Protpars protein parsimony method".

The main window displays a multiple sequence alignment of protein sequences. The sequences are listed on the left, including "Ov-TSP-1", "C. sinensis", "S. mansoni", "O. viverrini", and "S. japonicum". The alignment is shown in a grid format with columns representing positions. A scale bar at the top of the alignment area shows positions 80, 90, 100, 110, 120, and 130. The alignment is color-coded by amino acid type.

3. Run ClustalW

The screenshot shows the BioEdit Sequence Alignment Editor interface. The main window displays a sequence alignment of 8 total sequences. A ClustalW Options dialog box is open in the foreground, providing configuration settings for the ClustalW multiple alignment algorithm.

BioEdit Sequence Alignment Editor - [D:\OV_EXOSOME_Project\TSP2 project\TSP2_manuscript\input sequences.gb]

File Edit Sequence Alignment View Accessory Application RNA World Wide Web Options Window Help

8 total sequences

Mode: Select / Slide Selection: 0 Position: 1: Ov-TSP-14 Sequence Mask: None Numbering Mask: None Start ruler at: 1

ClustalW Options

ClustalW Multiple alignment

Reference:
Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994)
CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice.
Nucleic Acids Research, submitted, June 1994.

Full Multiple alignment

Calculate NJ Tree

FAST algorithm for guide tree

Bootstrap NJ Tree Number of bootstraps: 1000

Gap penalties: Blank=default

Pairwise alignments	Multiple alignment
Gap open	Gap open
Gap extend	Gap extend

Other Parameters:

Note: enter additional parameters as a single line.

Output Clustal format with Clustal consensus sequence generation

Additional Parameters for ClustalW:

```
****General settings:****  
/QUICKTREE :use FAST algorithm for the alignment guide tree  
/NEWTREE= :file for new guide tree  
/USETREE= :file for old guide tree  
/NEGATIVE :protein alignment with negative values in matrix
```

Run ClustalW View ClustalW Doc Cancel

Sequence alignment view showing positions 70-130. Sequences listed on the left include Ov-TSP-1, C. sinensis, S. mansoni, O. viverrini, C. sinensis, S. mansoni, S. japonicum, and O. viverrini.

Save as

BioEdit Sequence Alignment Editor - [Untitled1]

File Edit Sequence Alignment View Accessory Application RNA World Wide Web Options Window Help

8 total sequences

shade threshold 100%

Mode: Select / Slide Selection: 0 Position: Sequence Mask: None Numbering Mask: None Start ruler at: 1

Scroll speed slow fast

Ov-TSP-1
C. sinensis
S. mansoni C
O. viverrini
C. sinensis
S. mansoni C
S. japonicum
O. viverrini

---FLVVLGSEVVLGFSVYLSQPEPEAQDIIRASGHYVAQIALYALMGVGGITLITALFGCCGAYHESQILGAYFIIILIFTSCTGATLGYFEEEMQHVECCFEGVEEYSMLRDCRENPSPFMNIHRVLC
ATVVHMTLISVVLGFSVYLSQPEPEAQDIIRASGHYVAQIALYALMGVGGITLITALFGCCGAYHESQILGAYFIIILIFTSCTGATLGYFEEEMQHVECCFEGVEEYSMLRDCRENPSPFMNIHRVLC
---FLVVLGSEVVLGFSVYLSLEPEAQDIILASGQQAQIMYALMGIGGITLITALFGCCGAYHESQILGTYFTLILIFATCTGATLGYFEEEMKYVDDRREGIEEYSMLRDCRENVPVFMNISRQMLR
TFVVIACGVGLIVVGSIAEISLKT-----FGESNETSIQIIVIFILCLGCLTFILGFLGFCGACIKNVCMITYAILLGTAVACVCGIIGLILIRKIPGLVNHNEVLYTEYSANKDV----CKLINVLCSELK
SSRPLCCGGLIVVGSIAEISLKT-----YLSSEDACTAFVIFILAFGCFLTIVGSGFVFCGACKKRVCCITMYIIFVVFILGGVAAGIAGFVILKEDVKEYVDKILTCTYKTYTEEVS----RKLILILKQDLK
-ILVFACGLIIVVGSLSQAINN-----YSSGIDSNIKGLVIFILVLCFLFELGFLGFCGACTKNTCCILYAILLISMVAAEVGVGIAAAALREDVKTQFLSLRSSVSEYSKNPDI----KKFLIKQQEF
-ILVFACGLIIVVGSLSQAVNN-----YSTGIDSNIKGLVIFILVLCFLFELGFLGFCGACTKNTCCILYAILLISMVAAEVGVGIAAAALREDVKTQFLSLRNSVSEYSKNPDI----QNLFLIKQTEFK
-VVVICCGGLIIVVGSIAEISLKT-----YLSSEDACTAFVIFILAFGCFLTIVGSGFVFCGACKKRVCCITMYIIFVVFILGGVAAGIAGFVILKEDVKEYVDKILTCTYKTYNEEVS----RKLILILKQDLK

370 380 390 400 410 420 430 440 450 460 470 480 490

Windows taskbar: 16:21 26/3/2561

View alignments

1. File

2. Graphic view

The screenshot displays the BioEdit Sequence Alignment Editor interface. The 'File' menu is open, showing various options. The 'Graphic View' option is highlighted in blue. The main window shows a sequence alignment with a ruler at the top and a sequence viewer below. The sequence viewer displays three lines of amino acid sequences, with the first line being the reference sequence and the other two being aligned sequences. The alignment is shown with vertical bars indicating gaps and matches. The ruler at the top shows positions from 60 to 130. The sequence viewer shows the following sequences:

```
IIIVIFLCIGLTFLLIGFLGFGAATLNVMITVAILGVAVACVCGTIGLVLRKIPGLVNHNEVLYTEYSANKDVCRLI  
AEVIFLIAFGFLTVVGSFGFGACKLNVCCTMYIIFLVIFILGGTAAAGIAGFVLRSHVKEYVDKVIQTQYKTYNEEVS-KLI  
GLVIFLIILGFLFLLGFLGFGAATLNVCCTIMAILLSMVAAEIVGVIAAPVLRKIDVKTSEFVSVVRRNSVSEYSKNPDIQNLF  
GLVIFLIIVLIGFLFLLGFLGFGAATLNVCCTIIMAILLSMVAAEIVGAGIAAPVLRKIDVKTQFVLSVRRSSVSEYSKNPDI-RFL
```

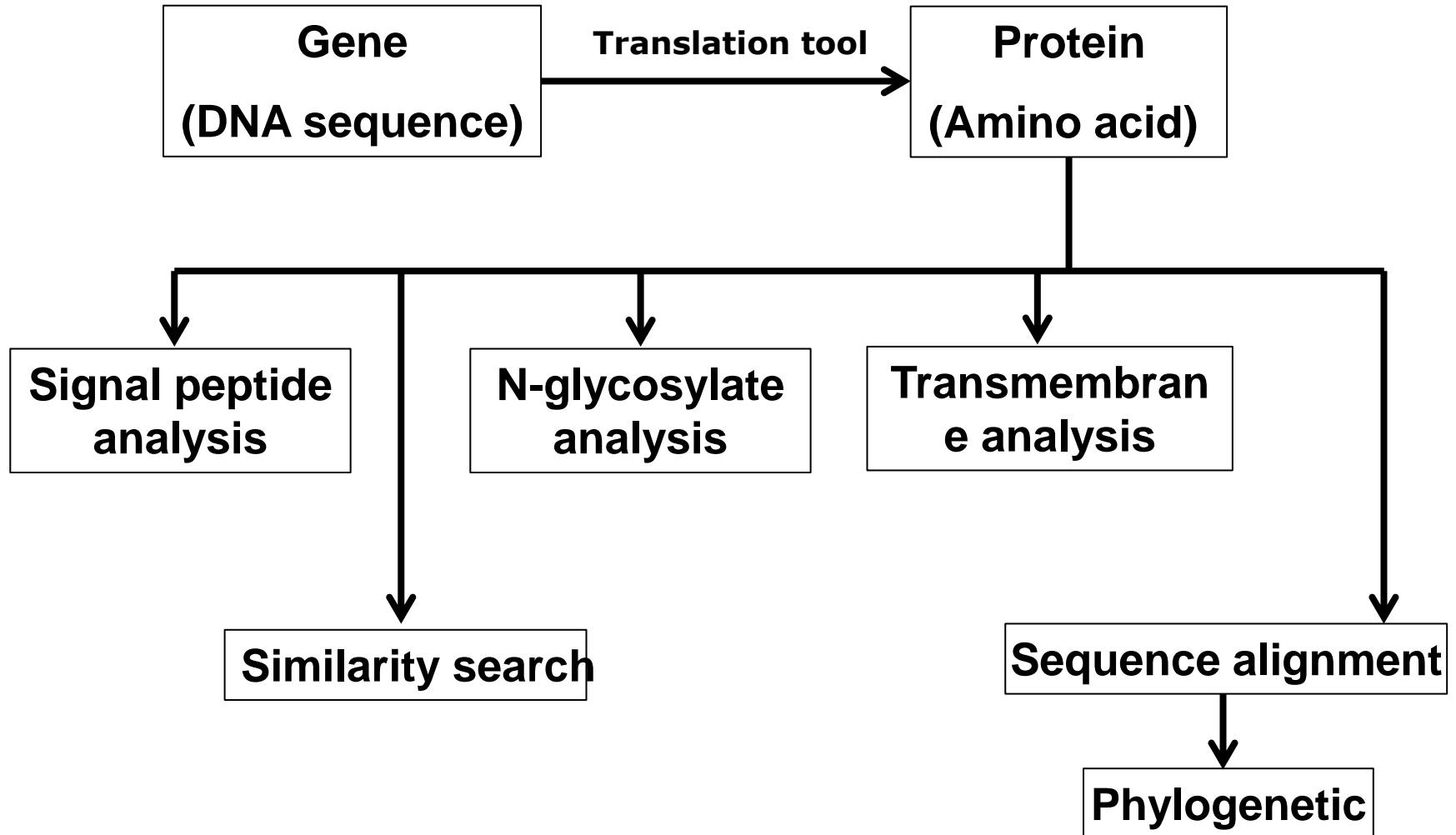
Font: Courier New Residues per row: 100 Normal Back Similar Back Identical Back Canvas Redraw
Size: 10 Characters in titles: 30 Font Font Titles AutoFit to Printer
Threshold (%) for shading: 50 Font Font Ruler Start numbers at 1
 Show main title Blocks of 10 residues start codon(s) allowed ATG Title relative to top
 Titles on left 3-letter translations 1-letter translations
 Titles on right numbers left of seqs numbers right of seqs Matrix BLOSUM62 Page 1
 Titles Bold Sequences Bold Show Ruler Page Height: 11 in.
 Titles Italicized Sequences Italicized Ruler Bold Page Width: 8.5 in.
 Titles Underlined Sequences Underlined Ruler in italics
 Id./Sim. Shading Sequences in color
 Id./Sim. Shading with color table Note: This is a print preview. It may copy to the clipboard slightly large. To expand view, press arrow (upper left)

```
      10      20      30      40      50      60      70      80      90
OV_TSP2A MVSLSGVRCLQIILVINTIETVIACGVGLIVVGSIAEVS LRTTEGSENETS IQITVIFITILGCLTFLIGFLGFCGACLRNVCYLITAYAIL
OV_TSP2B MVSLSGVRCLQIILVFNVIICGIALIVVGSIAEVLRTYISSEDAQLMAEVIITIAEGCLLIVVGSIFGFCGACLRNVCYLITAYITL
CD63_Sj MPTLSGVRCLQIILVFNILVFACGIALIVVGSISCVAVNNYSIGDINSIRGIVIFITILGCLFLLGLFLGFCGACTRNTCCCLITAYAIL
CD63_SM MPTLSGVRCLQIILVIFENILVFACGIALIVVGSISCVAVNNYSIGDINSIRGIVIFITILGCLFLLGLFLGFCGACTRNTCCCLITAYAIL

      110     120     130     140     150     160     170     180     190
OV_TSP2A CGIIGLVLRDRIPGLVNHENLEVIETSEYSSNEDVQRLLINVIQSELRCGGATGT--WAN-EGSEPESCRSP-EGVYTRDGGVFEVVEEPIQEN
OV_TSP2B AGIAGVLRDVKRYVDIVLTPQYKTYNEEVS--RRLIDLIQKLLCCGPDGT--WPPGLGIVPDSCRDS--SGLYTGCCSAALDRFIERN
CD63_Sj VGIAAAVLREDVRTSEVSVVRNSVSEYSRNPEDIQLEDRICDFRCCGSESSDYTSSGQSVPSGCDSDTGLAYQEGCSNRIIFAEERY
CD63_SM AGIAAVLREDVRTSEVLSVSEYSSVSEYSRNPDIRKFLDRICDFRCCGSESSDYTSSGQTPDSRNPNTAVVYSDDGCSNRFVISEEERY

      210     220
OV_TSP2A FALQLICMTFAICVVCALRGEFETV
OV_TSP2B FALQLIALLVEAVCVQAIQRGEDA--
CD63_Sj FALQLISIVFIVCIFAIRSGDGE--
CD63_SM FALQLLSIVFAVCVIFAIRSGDSD--
```

Flow chart for gene structure and function



Other tools on web

- Signal peptide analysis;

<https://services.healthtech.dtu.dk/service.php?SignalP-5.0>

- Transmembrane analysis

<https://services.healthtech.dtu.dk/service.php?TMHMM-2.0>

- EXPASY (Bioinformatics resource portal of the SIB Swiss Institute of Bioinformatics)

<https://www.expasy.org/>