



Mini Project: Comprehensive Sequence Analysis of the Human TNF Gene

Prepared by: **Chamss DAGHSNI**

Date: **23/08/2024**

Mini Project: Comprehensive Sequence Analysis of the Human TNF Gene

Objective:

To apply bioinformatics skills learned in Module 1 to download, analyze, and interpret the sequence of the human TNF gene, which encode a proinflammatory cytokine call TNF.

Project Overview:

In this mini project, I will perform a series of bioinformatics tasks using the human TNF gene as my sequence of interest. The project will guide me through downloading the sequence, translating it, finding ORFs, analyzing sequence composition, identifying transcription factor binding sites, searching for functional motifs, predicting coding/non-coding regions, and converting sequence file formats.

Task 1: Download a Biological Sequence from NCBI and View/Edit It

Objective:

Download the human TNF gene sequence and view it using BioEdit.

Instructions:

- Access the NCBI homepage at NCBI.
- Search for the human TNF gene using the term 'human TNF gene.'
- Locate the correct sequence record (e.g., 'Homo sapiens TNF').
- Download the sequence in FASTA format.
- Open the sequence in BioEdit and view/edit it.

Output:

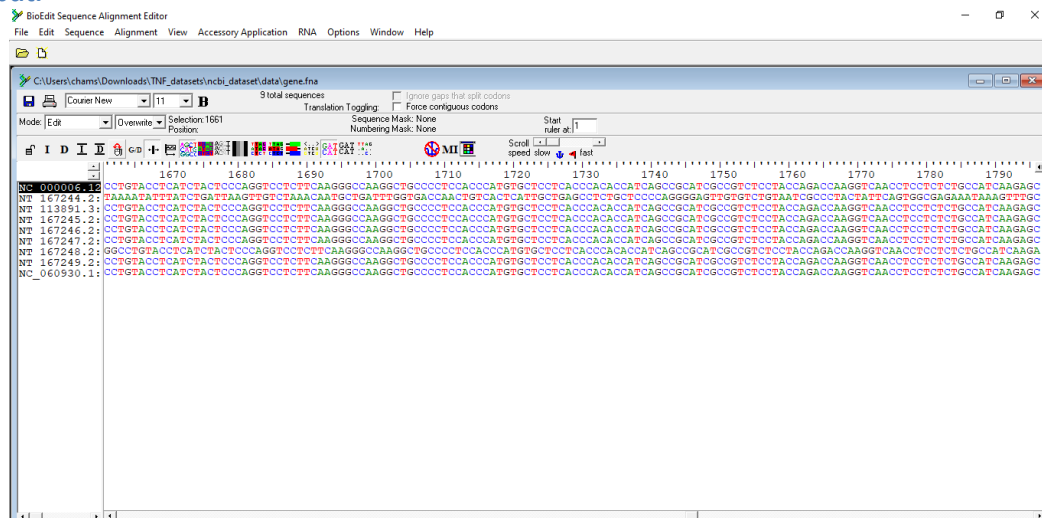


Figure 1: A screenshot of the human TNF gene sequence display in BioEdit

Task 2: Generate a Translation of a DNA or RNA Sequence into Amino Acids

Objective:

Translate the DNA sequence of the TNF gene into an amino acid sequence.

Instructions:

- Open the downloaded TNF gene sequence in BioEdit.
- Use the 'Translate' feature in BioEdit to generate the amino acid sequence.

Output:

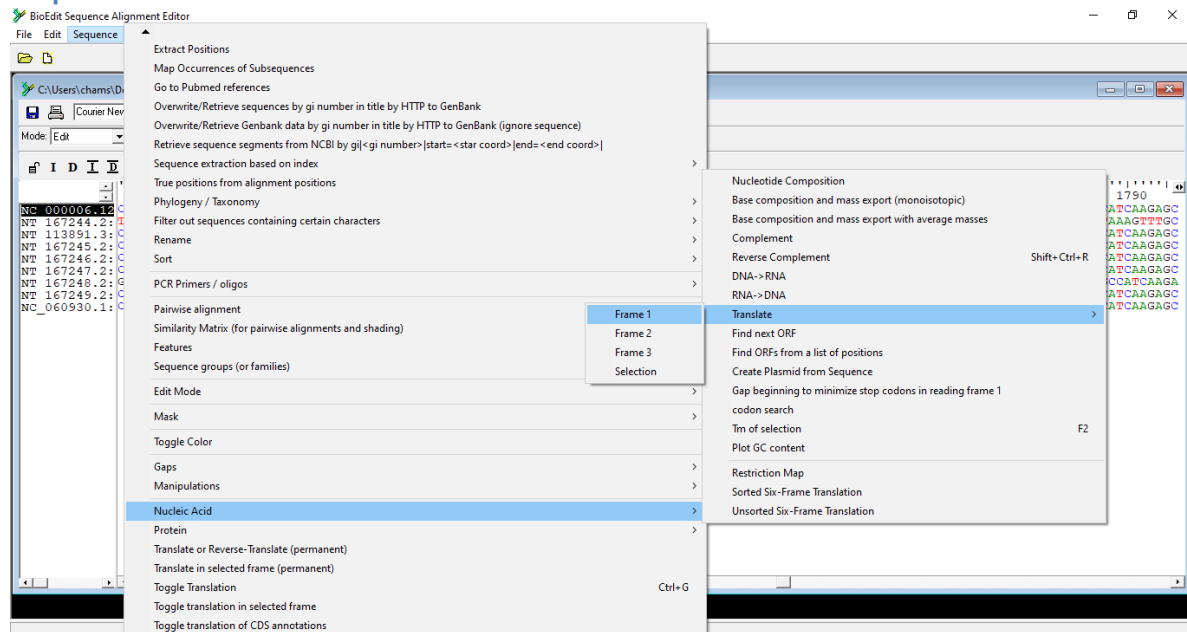


Figure 2: Access the Translate Feature

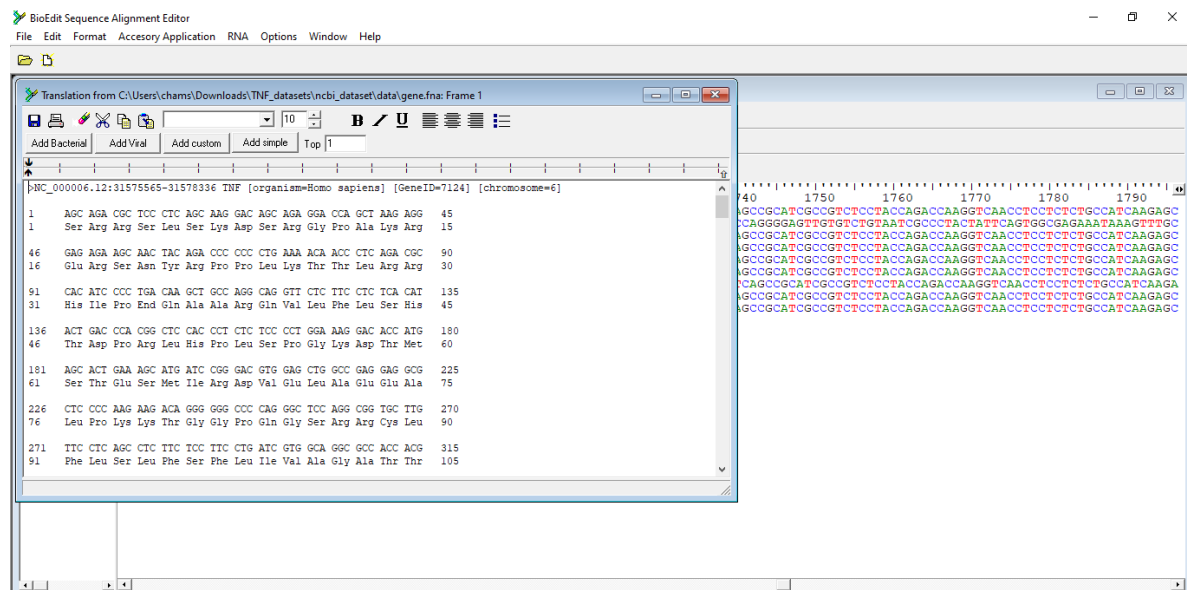


Figure 3: A screenshot of the amino acid generated in BioEdit

Task 3: Find ORFs (Open Reading Frames) in a DNA or RNA Sequence

Objective:

Identify the ORFs within the TNF gene sequence.

Instructions:

- Use BioEdit's ORF Finder tool to find ORFs in the TNF gene sequence.
- Record the start and stop positions, lengths, and protein translations of the ORFs.

Output:

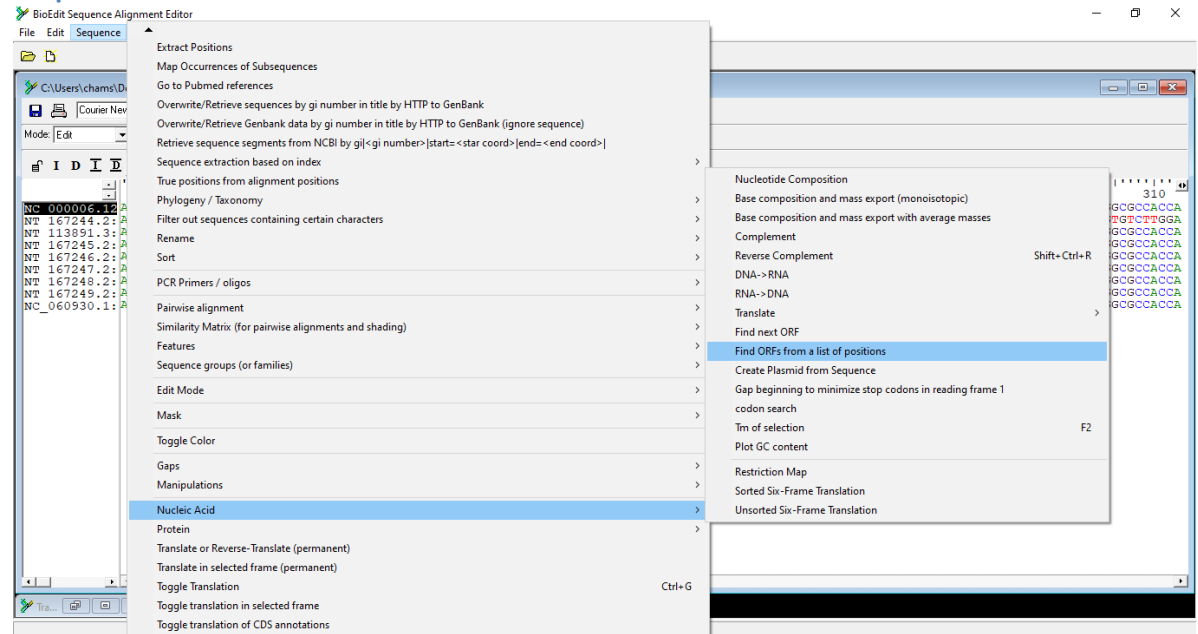


Figure 4: Access the ORF Finder Tool

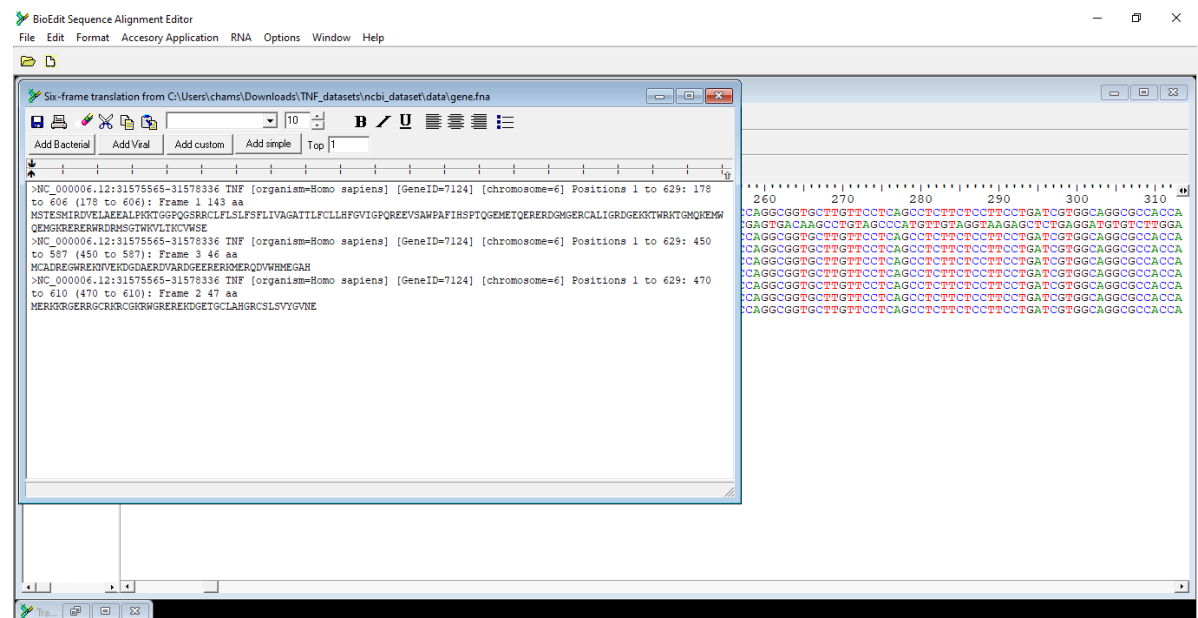


Figure 5: A screenshot of ORFs within the TNF gene sequence in BioEdit

Interpretation:

The ORF Finder tool in BioEdit identified three open reading frames (ORFs) in the TNF gene sequence.

The first ORF starts at position 178 and ends at position 606, producing a protein of 143 amino acids.

The second ORF begins at position 450 and ends at position 587, with a length of 46 amino acids.

The third ORF spans from position 470 to 610, encoding a protein of 47 amino acids.

These ORFs represent potential protein-coding regions within the TNF gene, with the first ORF likely corresponding to the main functional protein.

Task 4: Analyze Sequence Composition (Nucleotide or Amino Acid Frequencies)

Objective:

Analyze the nucleotide composition of the TNF gene sequence.

Instructions:

- Use BioEdit to analyze the sequence composition of the TNF gene.
- Calculate the frequencies of each nucleotide and the overall GC content.
- Interpret the results and save the analysis.

Output:

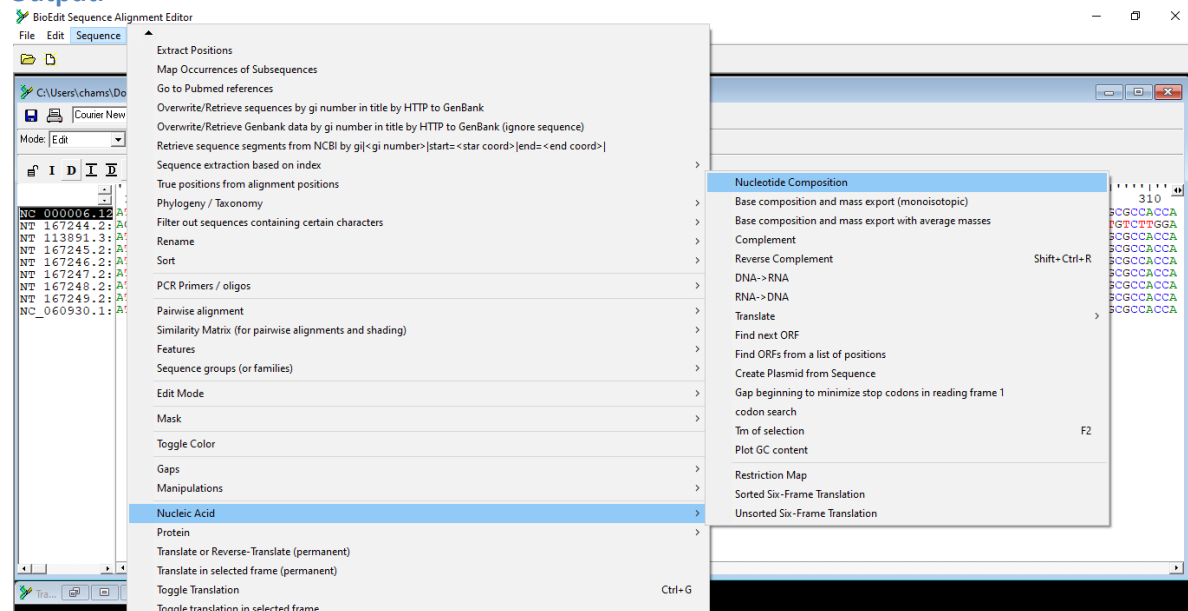


Figure 6: Access the nucleotide composition Tool

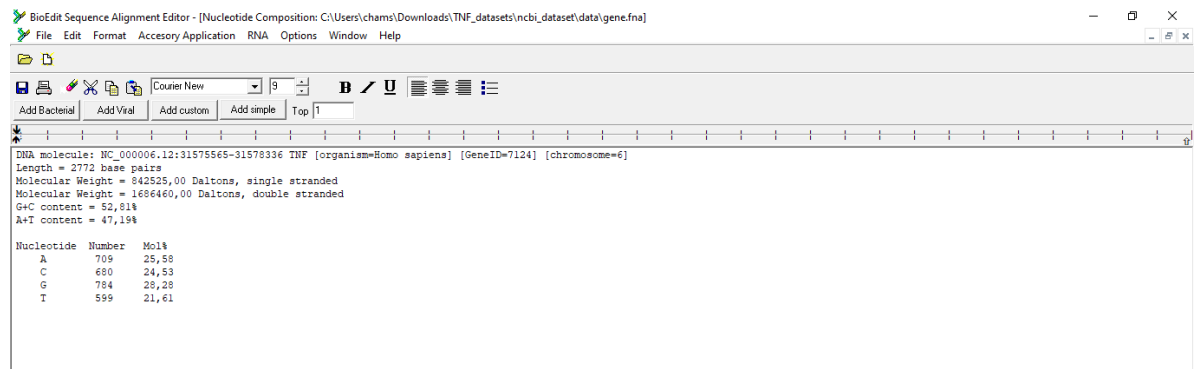


Figure 7: Screenshot of the nucleotide composition

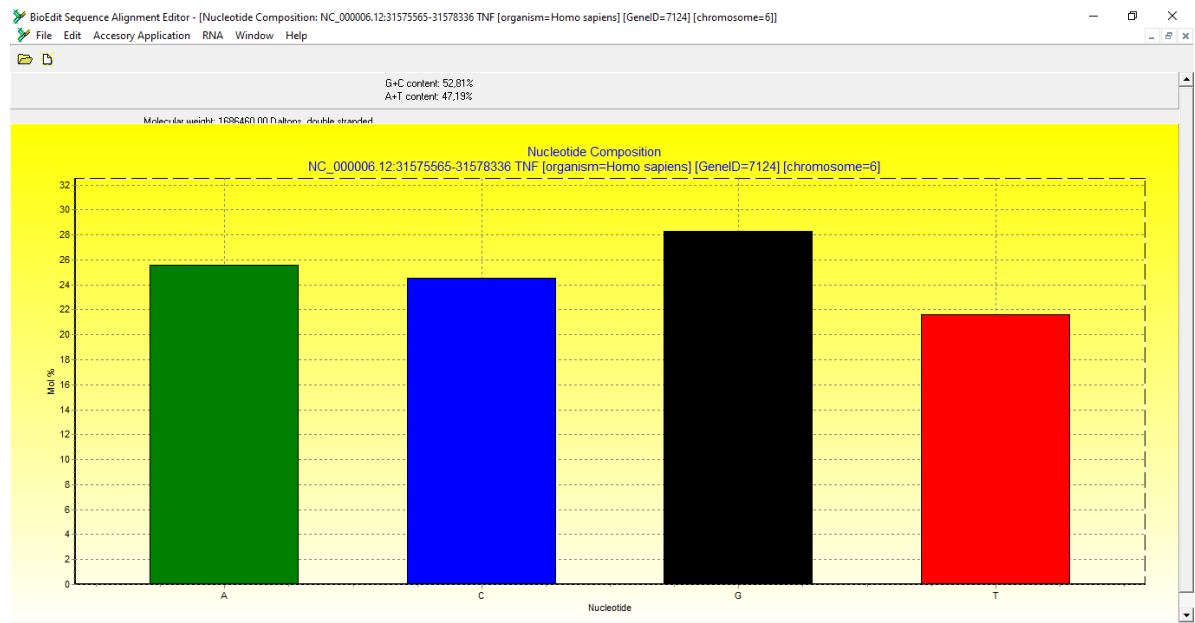


Figure 8: Graphical illustration of the nucleotide composition

Interpretation:

- The nucleotide composition analysis of the TNF gene sequence reveals the following distribution: Adenine (A) at 25.58%, Cytosine (C) at 24.53%, Guanine (G) at 28.28%, and Thymine (T) at 21.61%.
- The overall GC content of 52.81% indicates a relatively high proportion of guanine and cytosine bases, which is associated with increased stability of the DNA molecule.
- In contrast, the A+T content is lower at 47.19%, suggesting that the TNF gene may have a higher melting temperature and be more resistant to denaturation.
- This nucleotide balance could reflect the gene's evolutionary adaptation and functional requirements. The high GC content may also influence the gene's regulatory and structural features.

Task 5: Identify Transcription Factor Binding Sites Using the PROMO Tool

Objective:

Identify potential transcription factor binding sites in the TNF gene promoter region.

Instructions:

- Access the PROMO tool at PROMO.
- Select 'Homo sapiens' as the species.
- Input the promoter region of the TNF gene or use the entire gene sequence.
- Identify potential transcription factor binding sites.

Output:

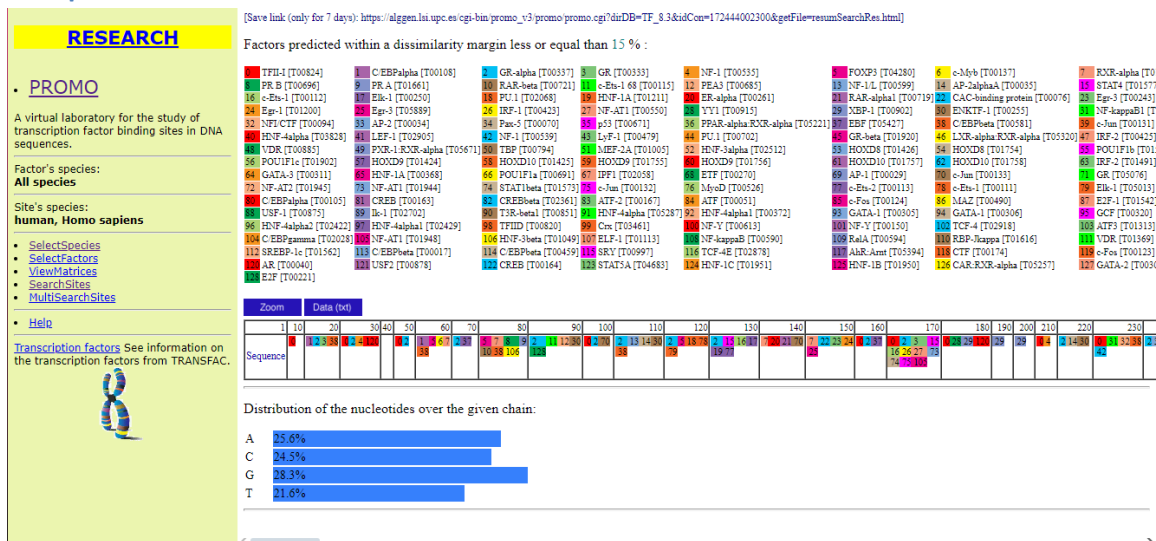


Figure 9: Transcription Factor Binding Sites Using the PROMO Tool

Interpretation:

The analysis identified numerous transcription factors that potentially bind to the promoter region of the TNF gene, including TFII-I, C/EBPalpha, GR, NF-1, FOXP3, and NF-kappaB. These transcription factors are involved in regulating immune responses, inflammation, and cellular stress, which are key processes influenced by TNF. The presence of binding sites for these factors indicates that TNF expression is tightly controlled by multiple regulatory mechanisms.

Task 6: Search for Functional Motifs in a Genome or Transcriptome Using MEME Suite

Objective:

Search for functional motifs in the TNF gene sequence using MEME Suite.

Instructions:

- Access the MEME Suite at MEME Suite.
- Upload the TNF gene sequence in FASTA format.
- Use the default settings to search for motifs.
- Interpret and save the results of the motif search.

Output:

The screenshot shows the MEME Suite Version 5.5.6 interface. On the left is a navigation menu with options like 'Motif Discovery', 'Motif Enrichment', 'Motif Scanning', 'Motif Comparison', 'Gene Regulation', 'Utilities', 'Manual', 'Guides & Tutorials', 'Sample Outputs', 'File Format Reference', 'Databases', 'Download & Install', 'Help', 'Alternate Servers', 'Authors & Citing', and 'Recent Jobs'. The main area is the 'Data Submission Form'. It includes a 'Select the motif discovery mode' section with 'Classic mode' selected. Below that is 'Select the sequence alphabet' with 'DNA, RNA or Protein' selected. The 'Input the primary sequences' section has a text area where a DNA sequence has been pasted. A 'Choose File' button is also visible.

Figure 10: Uploading the TNF gene sequence in the MEME Suite

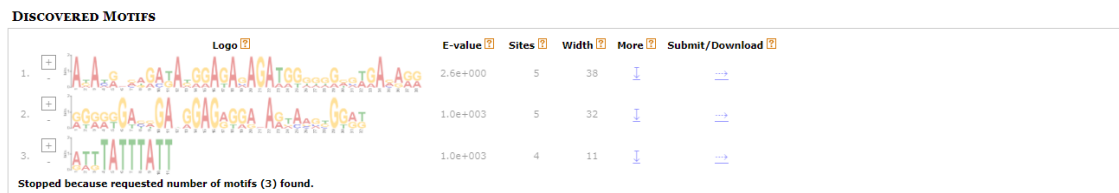


Figure 11: Discovered motifs



Figure 12: Details of the first discovered motif

Output:

GENSCAN Output

```
View gene model output: PS | PDF
GENSCAN 1.0      Date run: 23-Aug-124    Time: 17:39:50

Sequence /tmp/08_23_24-17:39:50.fasta : 2812 bp : 52.82% C+G : Isochore 3 (51 - 57 C+G%)

Parameter matrix: HumanIso.smat

Predicted genes/exons:

Gn.Ex Type S .Begin ...End .Len Fr Ph I/Ac Do/T CodRg P.... Tscr..
-----
1.01 Init +   218   403 186 1 0 94 105 207 0.702 22.03
1.02 Intr +  1010  1055  46 1 1 106 89  4 0.929  0.77
1.03 Intr +  1243  1290  48 2 0 140 82  25 0.987  6.24
1.04 Term +  1592  2013  422 0 2 132 55  518 0.985 48.73
1.05 P1yA +  2789  2794    6                      1.05

Suboptimal exons with probability > 1.000
```

Figure 15: GENSCAN Output

Interpretation:

The GENSCAN analysis indicates the presence of an initial exon (Init) starting at position 218 and ending at 403, followed by two introns (Intr) and a terminal exon (Term) from positions 1592 to 2013. The predicted exons are likely to encode significant portions of the TNF protein, while the introns suggest regions of non-coding sequences that may be involved in gene regulation or alternative splicing.

Task 8: Convert Between Sequence File Formats Using BioEdit (FASTA to PHYLIP)

Objective:

Convert the TNF gene sequence from FASTA format to PHYLIP format.

Instructions:

- Open the TNF gene sequence in BioEdit.
- Use the 'Save As...' feature to convert the file to PHYLIP format.
- Verify the conversion by opening the PHYLIP file in a text editor.

Output:

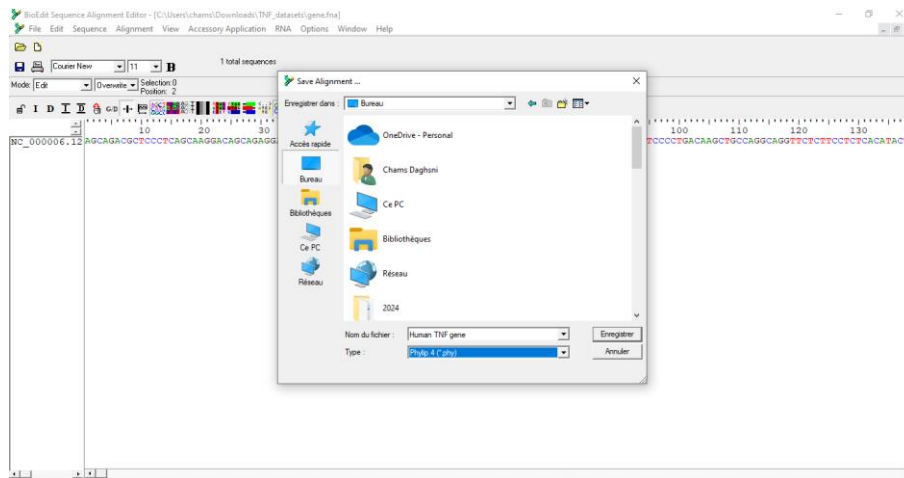


Figure 16: Converting the file to PHYLIP format in BioEdit

```
Human TNF gene - Bloc-notes
Fichier Edition Format Affichage Aide
1 2772
NC_000006.1 AGCAGACGCT CCTCAGCAA GGACAGCAGA GGACCACTA AGAGGGAGAG
AAGCAACTAC AGACCCCCC TGAAAAAC CCTCAGACG CACATCCCT
GACAAGCTGC CAGGCAGGTT CTCTCTCTCT CACATACTGA CCCACGGTC
CACCTCTCT CCTTGAAA GGACACCATG AGCACTGAAA GCATGATCCG
GGACGTGGAG CTGGCCGAGG AGGCGCTCCC CAAGAAGACA GGGGGCCCC
AGGGCTCCAG GCGTGCTTG TTCTCAGCC TTCTCTCTT CCTGATCTG
GCAGCGCCA CCACGCTCTT CTGCTGCTG CACTTTGGAG TGATCGGCC
CCAGAGGGA GAGTGAGTG CTGGCCAGC CTTCTCAC TCTCCACCC
AAGGGAAAT GGAGACCAA GAGAGGAGA GAGATGGAT GGTGAAAGA
TGTGCTGA TAGGAGGGA TGGAGAAA AAACGTGA GAAAGACGG
GATGCAGAA GAGATGTGC AAGAGATGG GAGAGAGAG AGAGAAAGAT
GGAGAGACG GATGCTCTGC ACATGGAAG TGCTACTAA GTGTGATATG
AGTGAATGAA TGAATGAAT AATGAACA GAGATATATA AATAGATAT
GGAGACAGAT GTGGGTGTG AGAAGAGAGA TGGGGGAAGA AACAGTGAT
ATGAATAAAG ATGGTGAGC AGAAGAGCG GGAATATGA CAGCTAAGGA
GAGAGATGG GGAGATAAG AGAGAAGAG ATAGGTGTC TGGCACACAG
AAGACACTCA GGAAGAGAG GTTGAATGC CTGGAAGGT AATACACAGA
TGAATGGAGA GAGAAAACCA GACACCTCAG GGCTAAGAG CAGGCCAGA
```

Figure 17: PHYLIP file in a text editor

