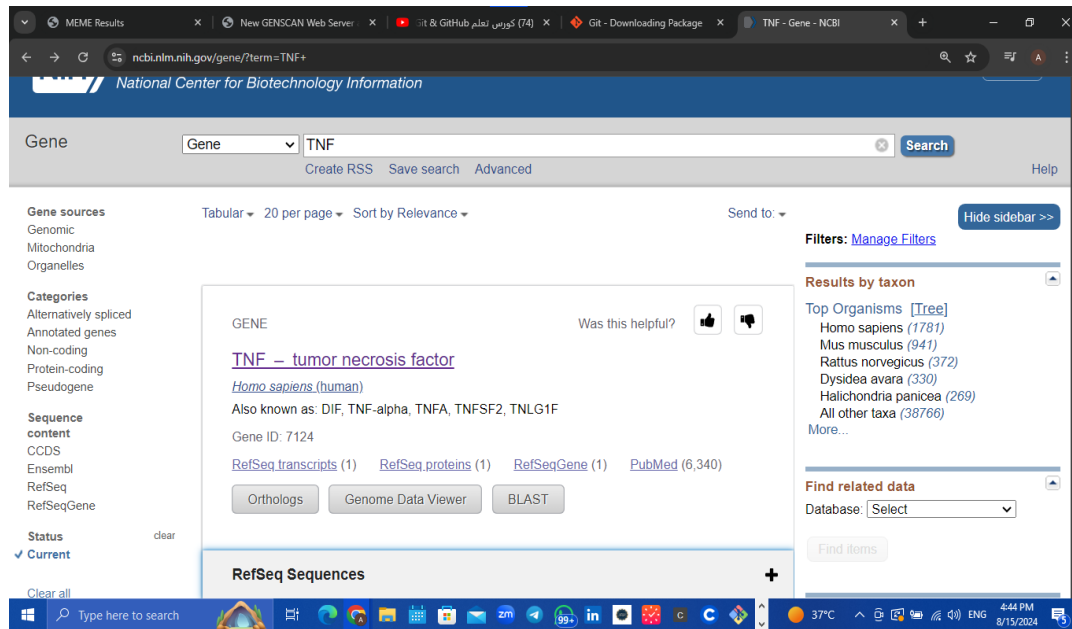




Comprehensive Sequence Analysis of the Human TNF Gene

Mini Project

Task 1: Download a TNF Gene Sequence from NCBI



The screenshot shows the NCBI Gene database interface for the TNF gene. The search bar at the top contains "TNF". The left sidebar lists various categories and sources. The main content area displays the gene name "TNF - tumor necrosis factor", its human origin, and its alternative names: DIF, TNF-alpha, TNFA, TNFSF2, and TNLG1F. The Gene ID is 7124. Below this, there are links to RefSeq transcripts (1), RefSeq proteins (1), RefSeqGene (1), and PubMed (6,340). A "RefSeq Sequences" section is visible at the bottom. The right sidebar shows "Results by taxon" with a list of top organisms: Homo sapiens (1781), Mus musculus (941), Rattus norvegicus (372), Dysidea avara (330), Halichondria panicea (269), and All other taxa (38766). There is also a "Find related data" section with a "Database" dropdown menu set to "Select".

OUTPUT TASK 1:

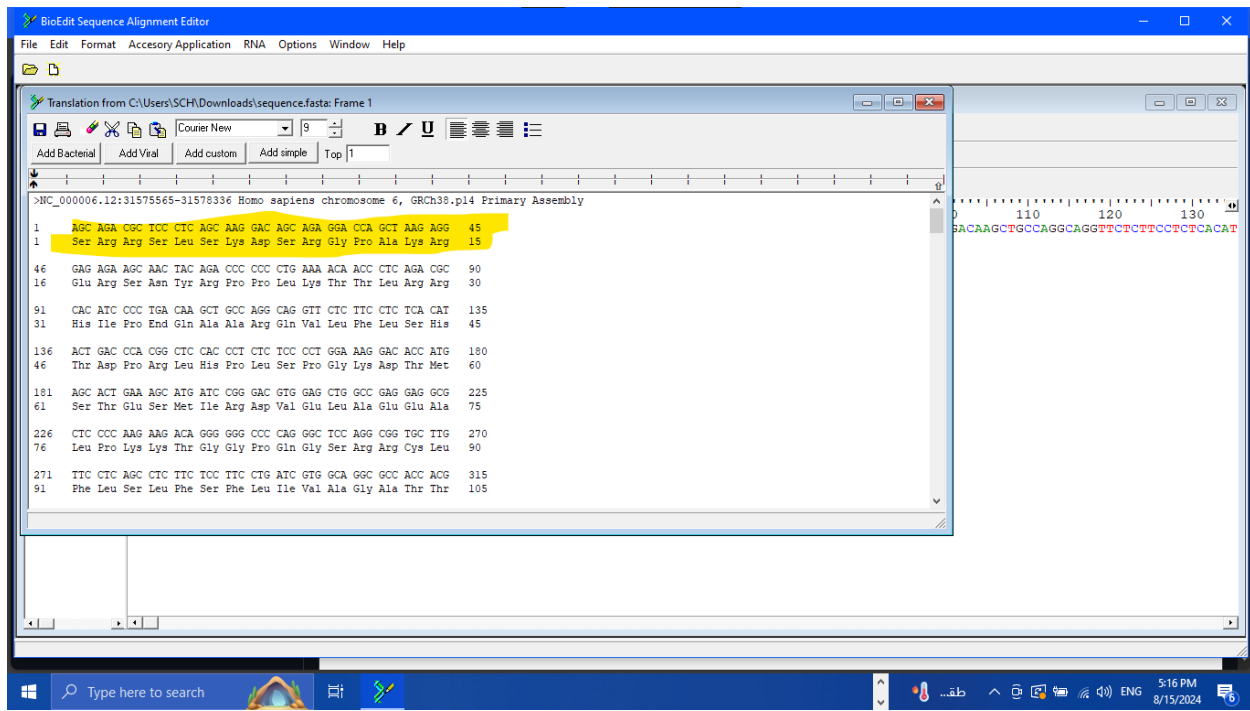
Download the TNF Gene Sequence as A Fasta_File

TASK 2: GENERATE A TRANSLATION OF A DNA OR RNA SEQUENCE INTO AMINO ACIDS

1- OPEN THE TNF_GENE_FASTA.FILE THAT IS DOWNLOADED IN THE FRIST TASK IN BIOEDIT

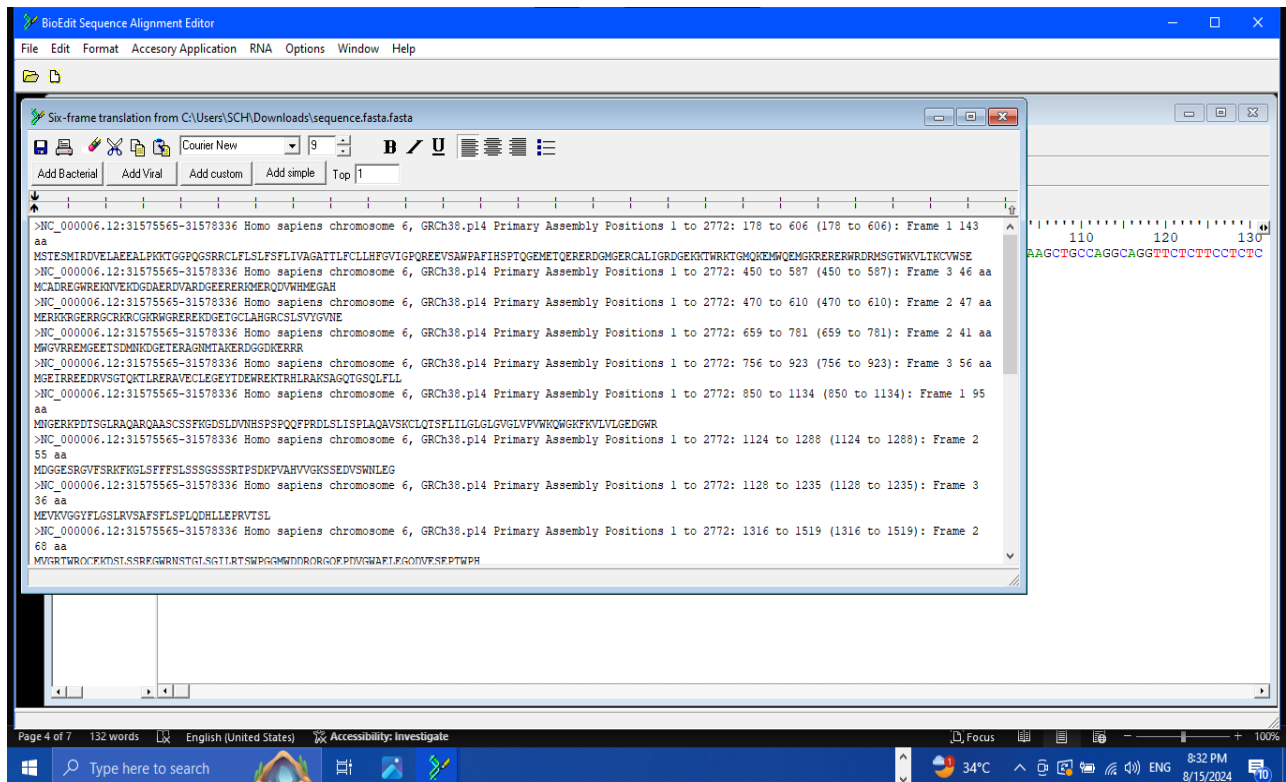
2- CHOOSE THE TRASLATE FROM NUCLIEC ACID

Output task 2 :



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

- 1- Open the TNF_Gene_fasta.file that is downloaded in the first task in Bioedit
- 2- choose the **find ORFS tool from list of positions** from the nucleic acid and select (1:2772 position)

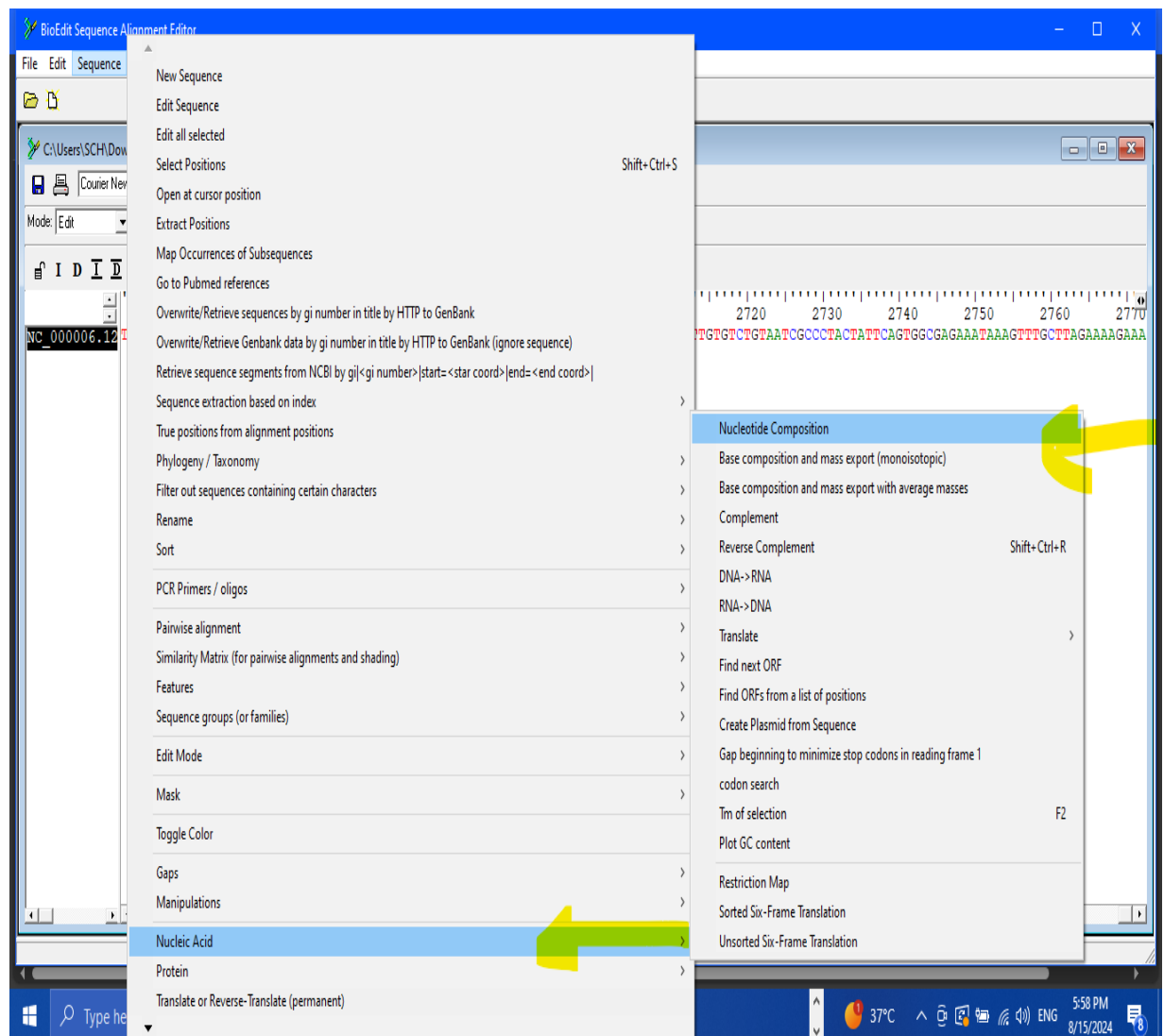


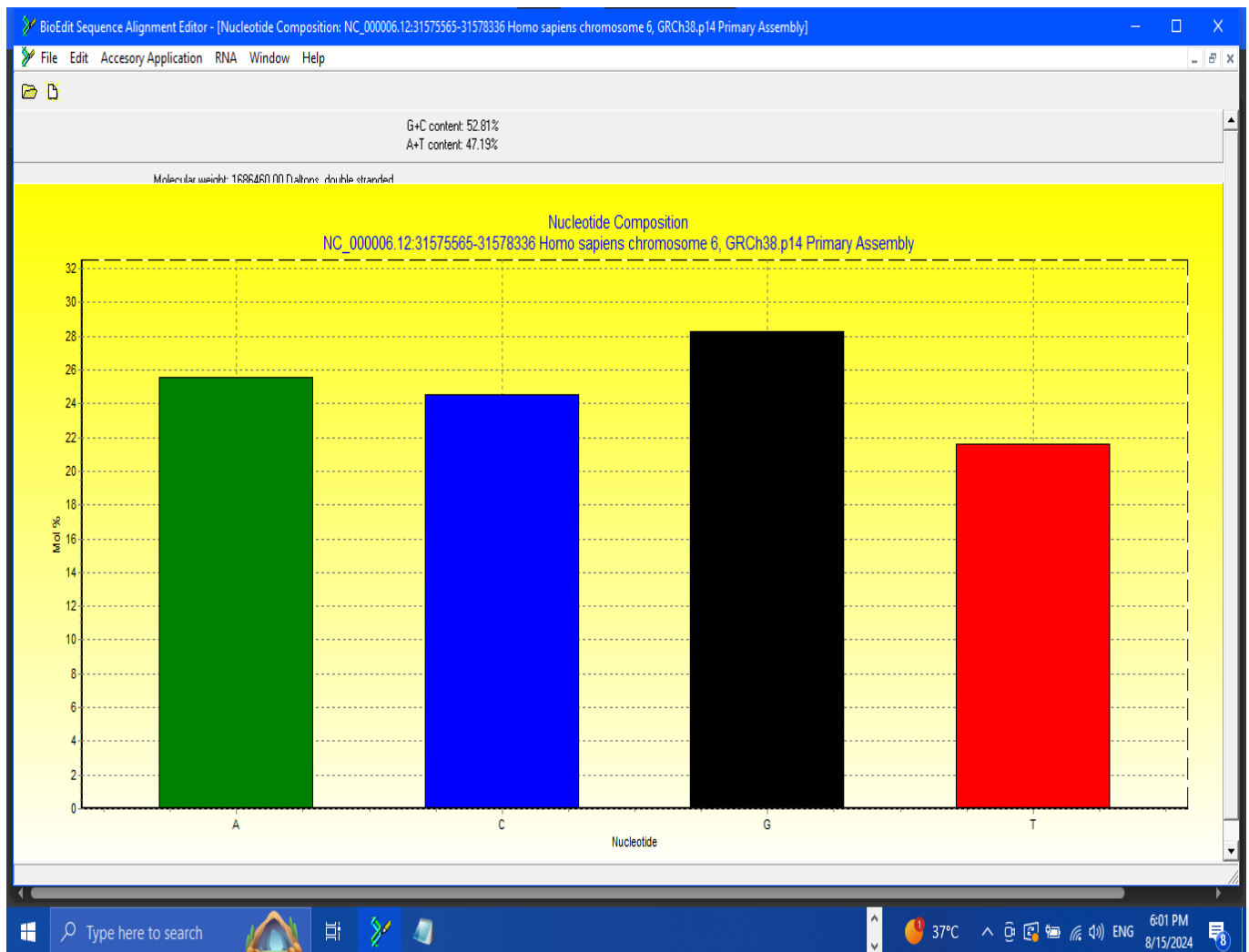
Output task 3 :

- 1_ understanding the protein-coding potential of the TNF_Gene
- 2_ identify the protein-coding regions of the TNF_Gene
- 3_ expect the changes in ORFs can lead to the gain or loss of protein function

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

- 1- open TNF_Gene_fasta.file on bioedit
- 2- choose from nucleic acid **Nucleotide composition tool**





Output task 4 :

- 1-percentage of the occurrence for each of the nucleotide A,T,C,G
- 2- (G-C) content is 52.81% it is make the sequence more stable

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

RESEARCH

PROMO

A virtual laboratory for the study of transcription factor binding sites in DNA sequences.

Factor's species:
human, Homo sapiens

Site's species:
human, Homo sapiens

- SelectSpecies
- SelectFactors
- ViewMatrices
- SearchSites
- MultiSearchSites

[Help](#)

[Transcription factors](#) See information on the transcription factors from TRANSFAC.



[Save link (only for 7 days): https://algen.lsi.upc.es/cgi-bin/promo_v3/promo/promo.cgi?dirDB=TF_8.3&idCon=172373441300&getFile=resumSearchRes.html]

Factors predicted within a dissimilarity margin less or equal than 15 % :

TFII-I [T00824]	GR-alpha [T00337]	FOXP3 [T04280]	c-Myb [T00137]	RXR-alpha [T01345]	PR-B [T00696]	PR-A [T01661]	RAR-beta [T00721]
PEA3 [T00685]	AP-2alpha [T00035]	STAT4 [T01577]	c-Ets-1 [T00112]	EBF-1 [T00250]	PU.1 [T02068]	ER-alpha [T00261]	RAR-alpha1 [T00719]
Egr-3 [T00243]	IRF-1 [T00423]	NF-AT1 [T00550]	YY1 [T00915]	XBP-1 [T00902]	ENKTF-1 [T00255]	NF-kappaB1 [T00593]	NF1-CTF [T00094]
Pax-5 [T00070]	p53 [T00671]	PPAR-alpha.RXR-alpha [T05221]	EBF [T05427]	C/EBPbeta [T00381]	HNF-4alpha [T03828]	LEF-1 [T02905]	NF-1 [T00539]
GR-beta [T01920]	VDR [T00885]	PXR-1.RXR-alpha [T05671]	TBP [T00794]	MEF-2A [T01005]	HNF-3alpha [T02512]	HOXD9 [T01424]	HOXD10 [T01425]
IRF-2 [T01491]	GATA-3 [T00311]	HNF-1A [T00368]	ETf [T00270]	AP-1 [T00029]	c-Jun [T00133]	GR [T05076]	NF-AT2 [T01945]
STAT1beta [T01573]	c-Ets-2 [T00113]	C/EBPalpha [T00105]	CREB [T00163]	ATF-2 [T00167]	ATF [T00051]	MAZ [T00490]	ELF-1 [T01542]
Ik-1 [T02702]	T3R-beta1 [T00851]	GATA-1 [T00306]	GCF [T00320]	TFIID [T00820]	NF-Y [T00150]	TCF-4 [T02918]	ELF3 [T01313]
NF-AT1 [T01948]	ELF-1 [T01113]	NF-kappaB [T00590]	RelA [T00594]	RBP-kappa [T01616]	SRY [T00997]	TCF-4E [T02878]	AhR-Arnt [T05394]
CTF [T00174]	c-Fos [T00123]	AR [T00040]	USF2 [T00878]	STAT5A [T04683]	HNF-1C [T01951]	HNF-1B [T01950]	GATA-2 [T00308]
E2F [T00221]							

Zoom Data (txt)

NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly

Distribution of the nucleotides over the given chain:

A	25.6%
C	24.5%
G	28.3%
T	21.6%

← → ↻ algen.lsi.upc.es/cgi-bin/promo_v3/promo/promo.cgi?dirDB=TF_8.3&idCon=172373441300&getFile=detailSearchRes.html ☆ A

NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly

1 25 50 75 100

AGCAGACGCTCCCTCAGCAAGGACAGCAGGACAGCTAAGAGGGAGAGAAGCAACTACAGACCCCTGAAACAAACCTCAGACGCCACATCCCTGACAAGCTGCCAGGCAGGTC



Outputs task 5 :

1- identify transcription factor binding sites in TNF_Gene sequence

2- Transcription Factor known to help us in regulating gene expression by binding to specific TNF_Gene sequences and either activating or repressing transcription

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

Suite

- 1- Search for functional motifs in the TNF gene sequence using MEME Suite
- 2-find 3 motifs and can know the beginning and ends each motif



Outputs task 6 :

- 1-Detected the motifs for binding a transcription factors
- 2-Motifs organize the regulation and the expression of a TNF_Gene

Task 7: Predict Coding/Non-Coding Regions in a Genome Using GENSCAN

Output:

1- KNOWN coding regions (exons) TO understande the function of different parts of the genome by its the beginning and ends and positions on sequence

2- insights into the regulation of gene expression from exons regions

```
GENSCAN 1.0      Date run: 15-Aug-124      Time: 11:33:09

Sequence /tmp/08_15_24-11:33:09.fasta : 2815 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 +   221   406 186 1 0  94 105 207 0.703 22.03
1.02 Intr +  1013  1058  46 1 1 106  89   4 0.929  0.77
1.03 Intr +  1246  1293  48 2 0 140  82  25 0.987  6.24
1.04 Term +  1595  2016  422 0 2 132  55 518 0.985 48.73
1.05 PlyA +  2792  2797   6                1.05

Suboptimal exons with probability > 1.000

Exnum Type S .Begin ...End .Len Fr Ph B/Ac Do/T CodRg P.... Tscr..
```

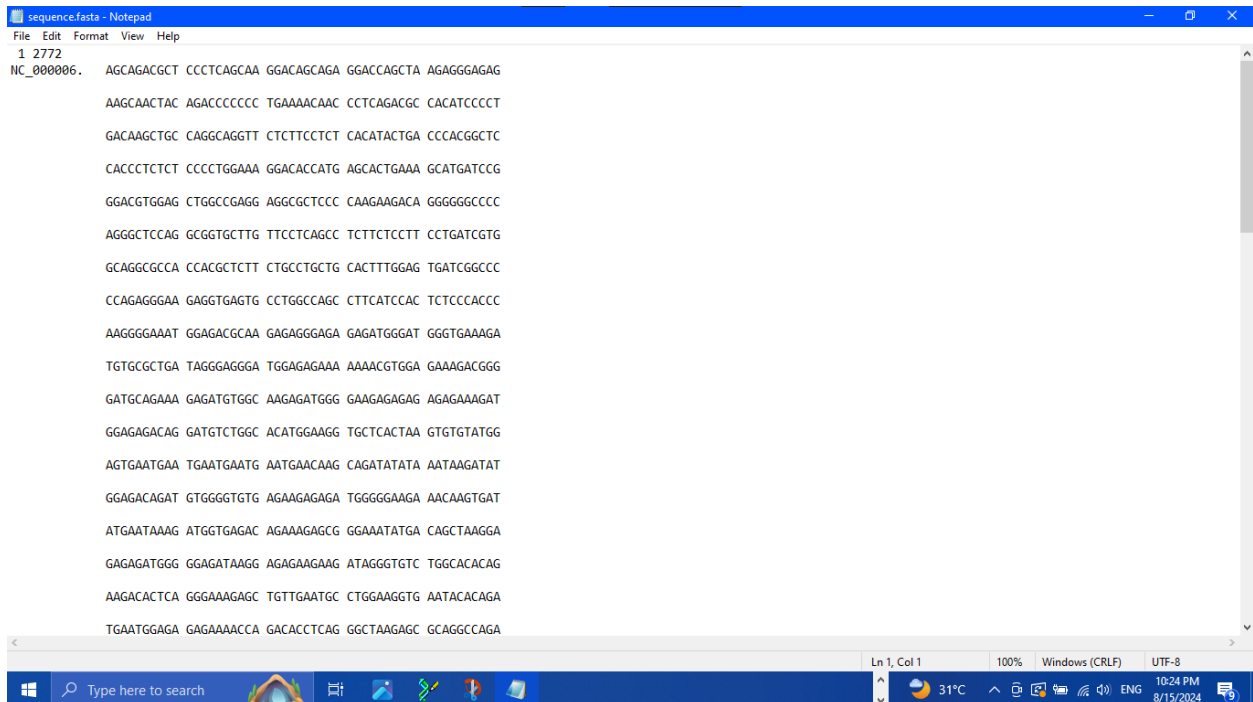
Task 8: Convert Between Sequence File Formats

Using BioEdit (FASTA to PHYLIP)

1- Open the *TNF_gene_fasta.file* in BioEdit.

2- Use the *Save As* from list of file

3- open the *PHYLIP* file in a text editor for prepare our data for phylogenetic analysis using the *PHYLIP* software



```
sequence.fasta - Notepad
File Edit Format View Help
1 2772
NC_000006.  AGCAGACGCT  CCTCAGCAA  GGACAGCAGA  GGACAGCTA  AGAGGGAGAG
            AAGCAACTAC  AGACCCCCC  TGAAAAAAC  CCTCAGACGC  CACATCCCCT
            GACAAGCTGC  CAGGCAGGTT  CTCTTCTCT  CACATACTGA  CCCACGGCTC
            CACCTCTCT  CCCCTGAAA  GGACACCATG  AGCACTGAAA  GCATGATCCG
            GGACGTGGAG  CTGGCCGAGG  AGGCGCTCCC  CAAGAAGACA  GGGGGGCCCC
            AGGGCTCCAG  GCGGTGCTTG  TTCCTCAGCC  TCTTCTCCTT  CCTGATCGTG
            GCAGGCGCCA  CCACGCTCTT  CTGCCTGCTG  CACTTTGGAG  TGATCGGCCC
            CCAGAGGGAA  GAGGTGAGTG  CCTGGCCAGC  CTTCATCCAC  TCTCCCACCC
            AAGGGGAAAT  GGAGACGCAA  GAGAGGGAGA  GAGATGGGAT  GGTGAAAGA
            TGTGCGCTGA  TAGGGAGGGA  TGGAGAGAAA  AAAACGTGGA  GAAAGACGGG
            GATGCAGAAA  GAGATGTGGC  AAGAGATGGG  GAAGAGAGAG  AGAGAAAGAT
            GGAGAGACAG  GATGTCTGGC  ACATGGAAGG  TGCTCACTAA  GTGTGTATGG
            AGTGAATGAA  TGAATGAATG  AATGAACAAG  CAGATATATA  AATAAGATAT
            GGAGACAGAT  GTGGGTGTG  AGAAGAGAGA  TGGGGAAGA  AACAAGTGAT
            ATGAATAAAG  ATGGTGAGAC  AGAAAGAGCG  GGAATATGA  CAGCTAAGGA
            GAGAGATGGG  GGAGATAAGG  AGAGAAGAAG  ATAGGGTGTG  TGGCACACAG
            AAGACACTCA  GGGAAAGAGC  TGTGTAATGC  CTGGAAGGTG  AATACACAGA
            TGAATGAGA  GAGAAAACCA  GACACCTCAG  GCCTAAGAGC  GCAGGCCAGA
```