

BIOINFORMATICS MINI PROJECT: Comprehensive Sequence Analysis of the Human TNF Gene

TASK 1: Download the human TNF gene sequence from NCBI and view/edit it

Tool(s) used: NCBI and BioEdit

Output:

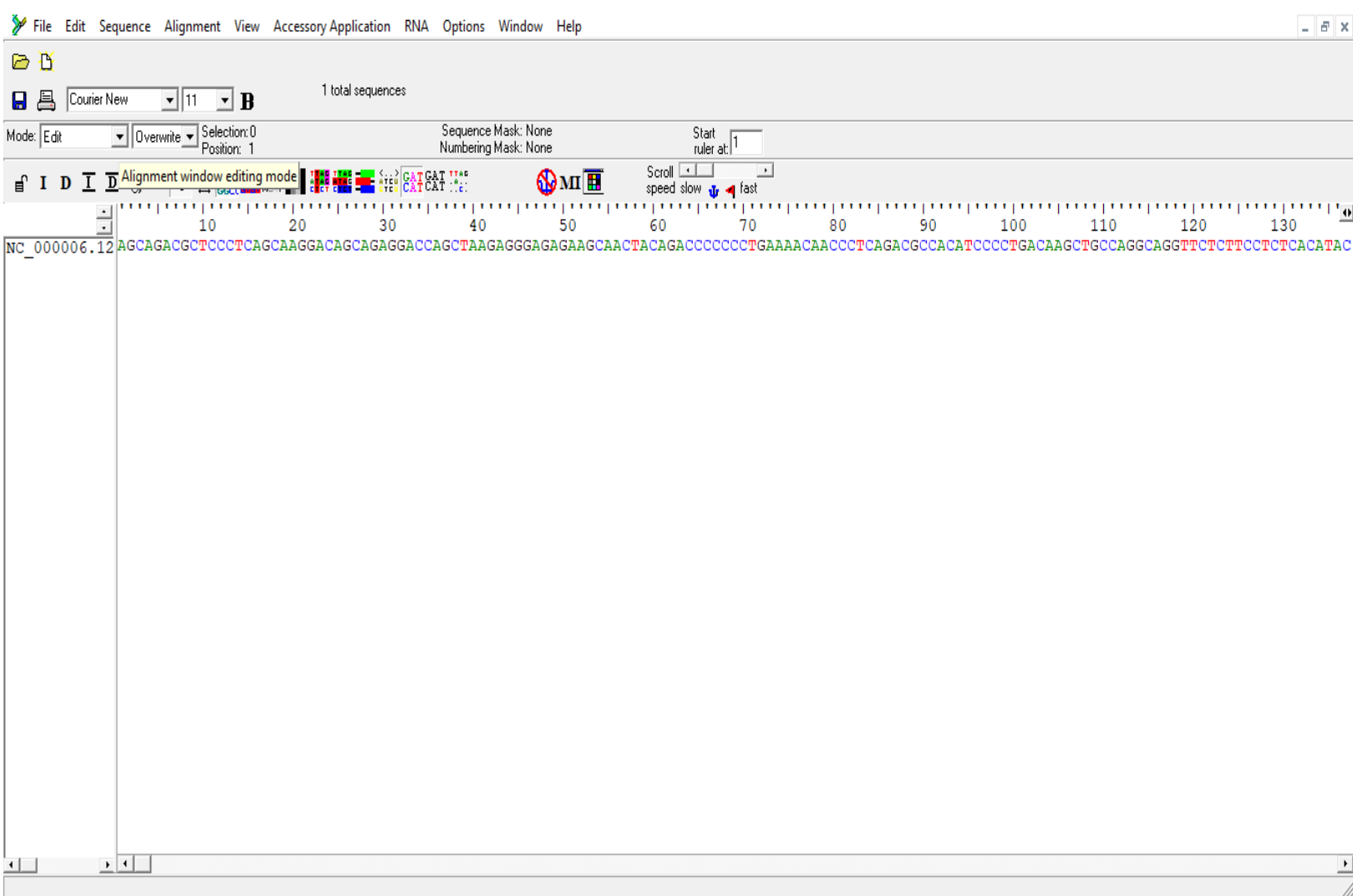


Figure 1: Human TNF gene sequence on BioEdit

TASK 2: Translate the DNA sequence of the TNF gene into an amino acid sequence

Tool(s) used: BioEdit

Output:

```
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly
1   AGC AGA CGC TCC CTC AGC AAG GAC AGC AGA GGA CCA GCT AAG AGG   45
1   Ser Arg Arg Ser Leu Ser Lys Asp Ser Arg Gly Pro Ala Lys Arg   15
46  GAG AGA AGC AAC TAC AGA CCC CCC CTG AAA ACA ACC CTC AGA CGC   90
16  Glu Arg Ser Asn Tyr Arg Pro Pro Leu Lys Thr Thr Leu Arg Arg   30
91  CAC ATC CCC TGA CAA GCT GCC AGG CAG GTT CTC TTC CTC TCA CAT   135
31  His Ile Pro End Gln Ala Ala Arg Gln Val Leu Phe Leu Ser His   45
136 ACT GAC CCA CGG CTC CAC CCT CTC TCC CCT GGA AAG GAC ACC ATG   180
46  Thr Asp Pro Arg Leu His Pro Leu Ser Pro Gly Lys Asp Thr Met   60
181 AGC ACT GAA AGC ATG ATC CGG GAC GTG GAG CTG GCC GAG GAG GCG   225
61  Ser Thr Glu Ser Met Ile Arg Asp Val Glu Leu Ala Glu Glu Ala   75
226 CTC CCC AAG AAG ACA GGG GGG CCC CAG GGC TCC AGG CGG TGC TTG   270
76  Leu Pro Lys Lys Thr Gly Gly Pro Gln Gly Ser Arg Arg Cys Leu   90
271 TTC CTC AGC CTC TTC TCC TTC CTG ATC GTG GCA GGC GCC ACC ACG   315
91  Phe Leu Ser Leu Phe Ser Phe Leu Ile Val Ala Gly Ala Thr Thr   105
316 CTC TTC TGC CTG CTG CAC TTT GGA GTG ATC GGC CCC CAG AGG GAA   360
106 Leu Phe Cys Leu Leu His Phe Gly Val Ile Gly Pro Gln Arg Glu   120
361 GAG GTG AGT GCC TGG CCA GCC TTC ATC CAC TCT CCC ACC CAA GGG   405
121 Glu Val Ser Ala Trp Pro Ala Phe Ile His Ser Pro Thr Gln Gly   135
406 GAA ATG GAG ACG CAA GAG AGG GAG AGA GAT GGG ATG GGT GAA AGA   450
136 Glu Met Glu Thr Gln Glu Arg Glu Arg Asp Gly Met Gly Glu Arg   150
451 TGT GCG CTG ATA GGG AGG GAT GGA GAG AAA AAA ACG TGG AGA AAG   495
```

Figure 2: Translation of the human TNF gene sequence into amino acid sequence

TASK 3: Identify the ORFs (Open Reading Frames) within the TNF gene sequence

Tool(s) used: BioEdit

Output:

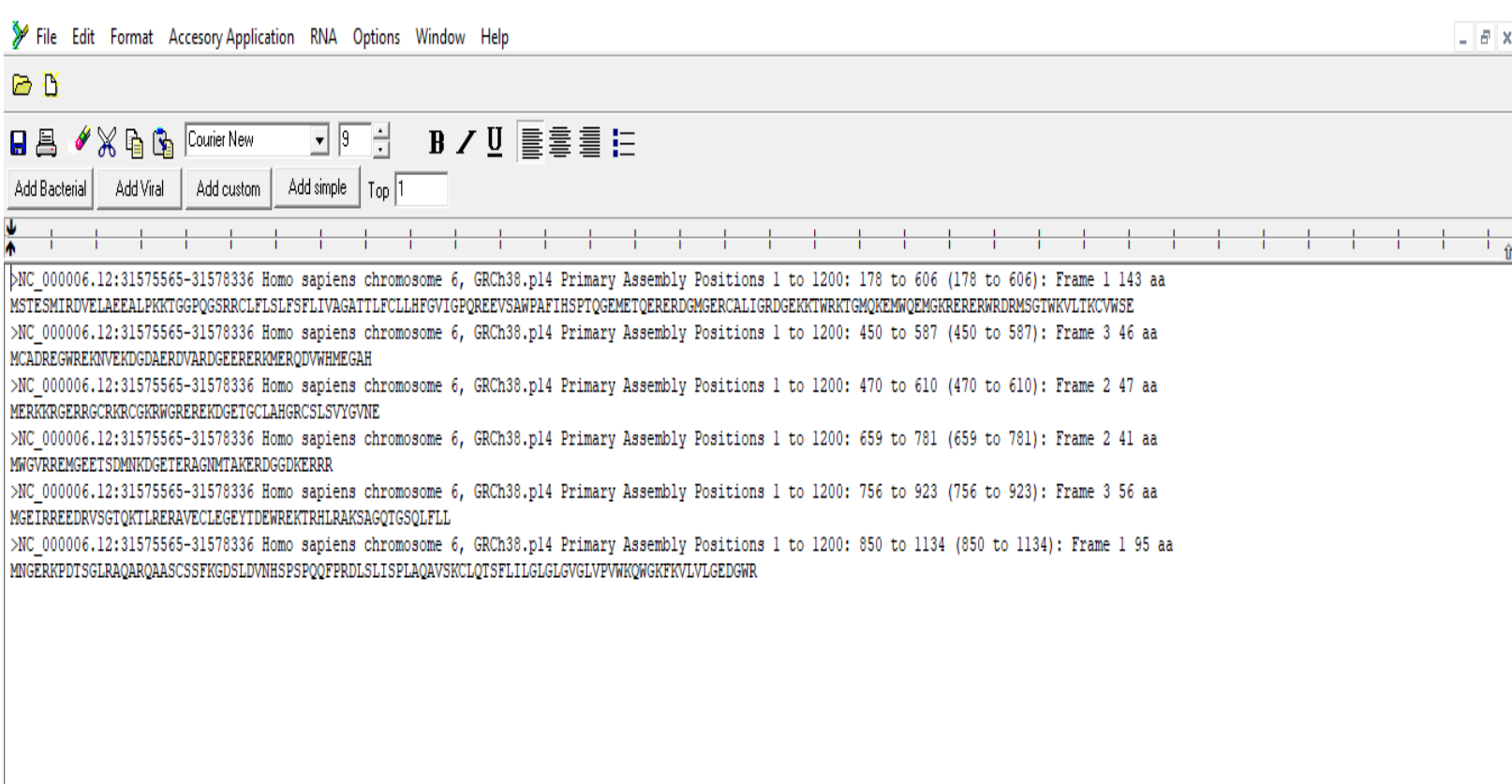


Figure 3: Identification of ORFs in the TNF gene sequence

Result interpretation:

- Figure 3 displays 6 open reading frames (ORFs) ranging from positions 1 to 1200.
- Each ORF is annotated with its start and stop positions, frame designation, and corresponding protein translations.
- The first ORF spans positions 178 to 606, producing a protein of 143 amino acids. Below, the amino acid sequence translation for that specific ORF is shown.
- Identifying ORFs is crucial because it helps researchers predict the locations of genes within a DNA sequence, including those that may play roles in biological processes and disease development.

Task 4: Analyze the nucleotide composition of the TNF gene sequence

Tool(s) used: BioEdit

Output:

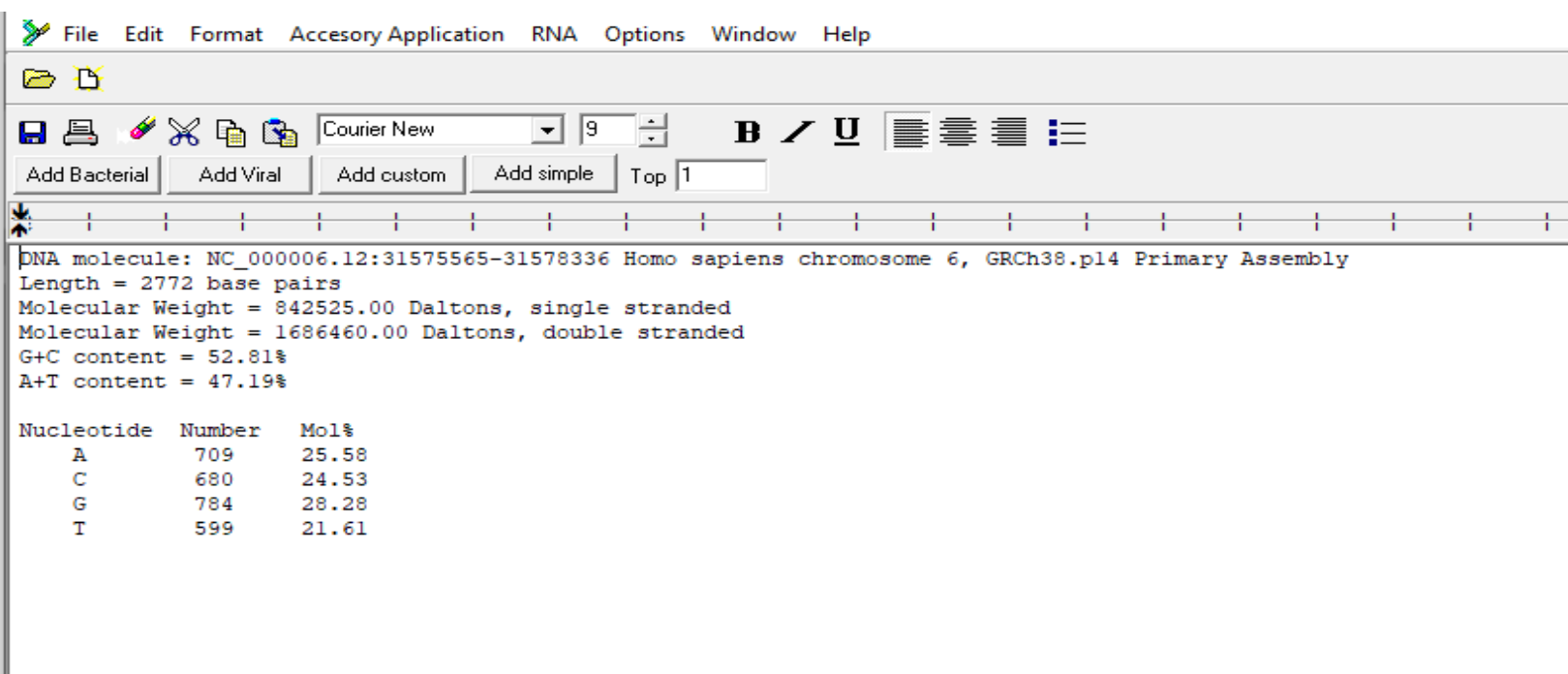


Figure 4: Analysis of nucleotide composition in the TNF gene sequence

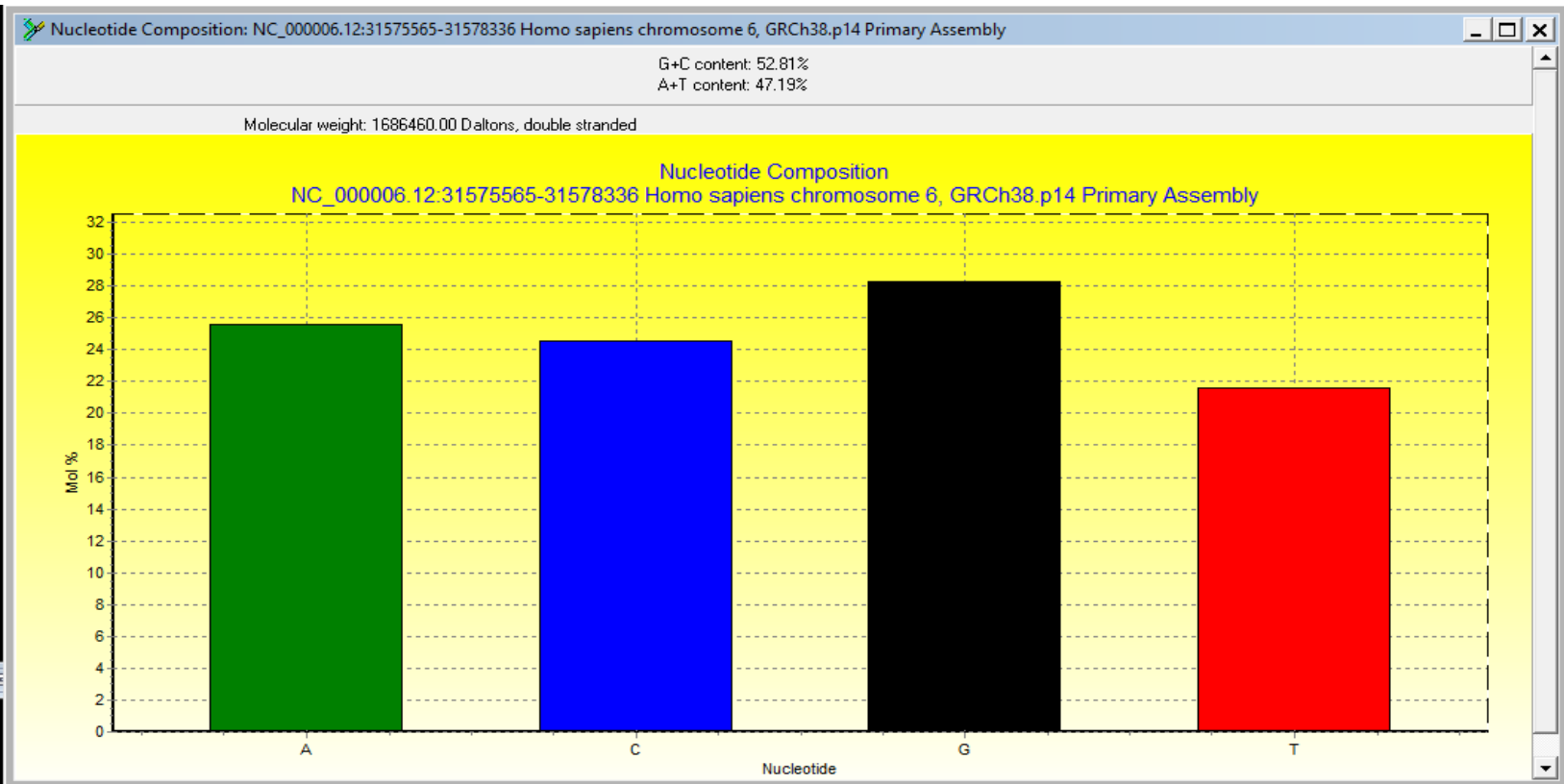


Figure 5: Analysis of nucleotide composition in the TNF gene sequence (Bar graph)

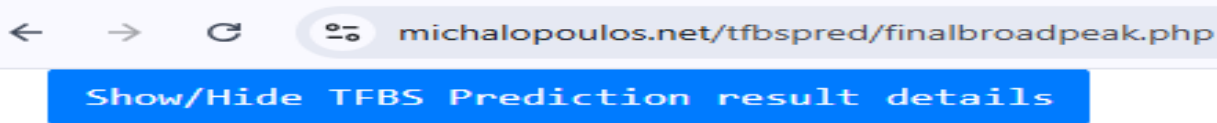
Result interpretation:

- Figures 4 and 5 illustrate the nucleotide frequency, molecular weight, and GC/AT content of the TNF gene sequence.
- The observed nucleotide counts are 709 adenine (A), 680 cytosine (C), 784 guanine (G), and 599 thymine (T).
- The sequence has an overall GC content of 52.81%.
- This value is particularly relevant because GC content influences the thermal stability of DNA, with higher GC levels generally conferring greater stability.

Task 5: Identify potential transcription factor binding sites in the TNF gene promoter region

Tool(s) used: TFBSPred

Output:



EGR1

hg38	31575388	TCCCCGCCCCCGCG	31575401
		. . .	
mm10	35202179	CCTCTGCCCCCGCG	35202166

ELF5

hg38	31575452	TGGAGGAAGCG	31575442
mm10	35202117	TGGAGGAAGCG	35202127

ERG

hg38	31575451	GGAGGAAGCG	31575442
mm10	35202118	GGAGGAAGCG	35202127

ETS1

hg38	31575451	GGAGGAAGCG	31575442
mm10	35202118	GGAGGAAGCG	35202127

Figure 6: Identification of potential transcription factor binding sites in the TNF gene sequence

Result interpretation:

- Figure 6 shows the predicted transcription factor binding sites in the TNF gene sequence using TFBSPred. Binding sites were found for factors such as EGR1, ELF5, ERG, and ETS1, with additional factors like PATZ1 - although not included in this figure.
- Each of these sites could contribute to the regulation of TNF gene expression by allowing specific transcription factors to bind and control activity. These predictions highlight possible regulatory elements that may influence TNF function in health and disease.

Task 6: Search for functional motifs in the TNF gene sequence using MEME Suite

Tool(s) used: MEME Suite

Output:

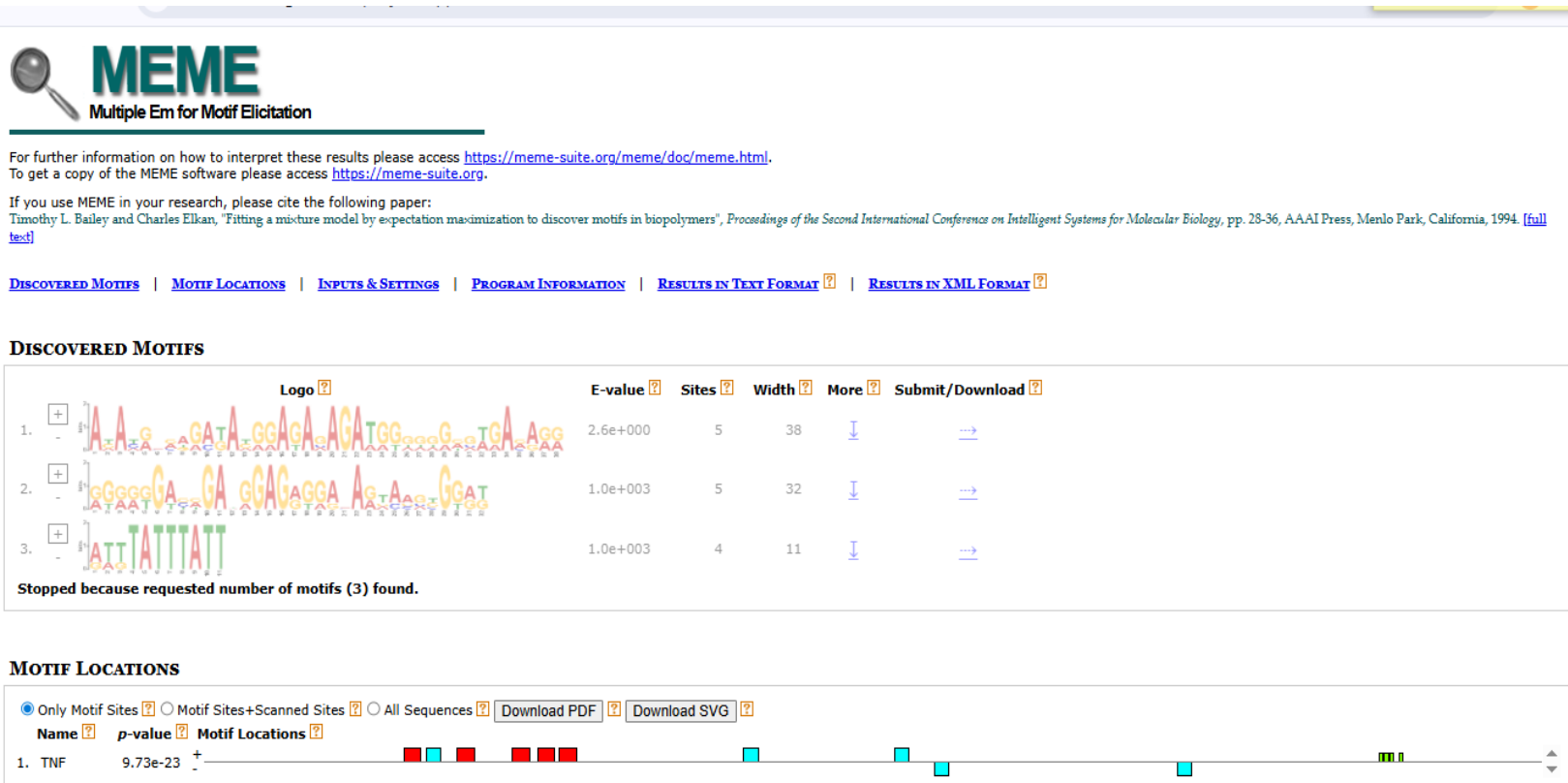


Figure 7: Identification of functional motifs in the TNF gene sequence

Result interpretation:

- Figure 7 shows three discovered motifs, along with their E-values, sites, lengths, and positions in the TNF gene sequence.
- The E-value indicates how statistically significant a motif is. Lower E-values indicate greater significance, while values above 0.05 are usually considered not significant and are shown in grey by MEME.
- In this result, all three motifs have E-values above 0.05, so they are non-significant and displayed in grey.

Task 7: Predict the coding and non-coding regions within the TNF gene sequence

Tool(s) used: GENSCAN

Output:

GENSCAN Output

View gene model output: [PS](#) | [PDF](#)

GENSCAN 1.0 Date run: 6-Sep-125 Time: 17:00:42

Sequence /tmp/09_06_25-17:00:42.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

Figure 8: Identification of coding and non-coding regions within the TNF gene sequence

Result interpretation:

- The output in figure 8 shows details for each predicted gene/exon, including sequence length, reading frame, type, strand orientation, and several specific scores (I/AC, Do/T, etc).
- In total, four exons were predicted: two internal (Intr), one initial (Init), one terminal (Term), followed by one predicted polyA site (PlyA). All of these are located on the positive (+) strand, as indicated in the third column (S); and most show high probability values.
- It should be noted that predicted exons with higher probabilities are generally more reliable than those with lower ones.

Task 8: Convert the TNF gene sequence from FASTA format to PHYLIP format

Tool(s) used: BioEdit

Output: Included within the upload