**Mini Project: Comprehensive Sequence Analysis of the Human TNF Gene**

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

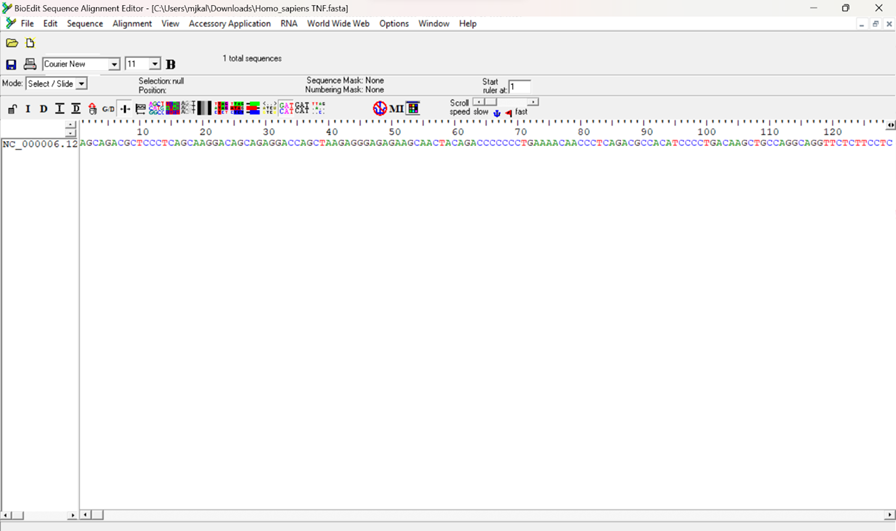


Figure1

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

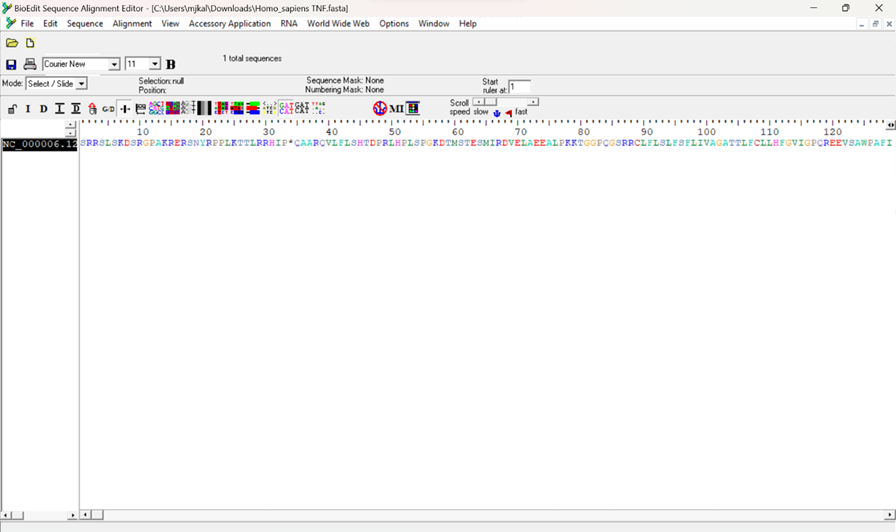


Figure 2

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

* In Figure 3 the start codon (ATG) is present.

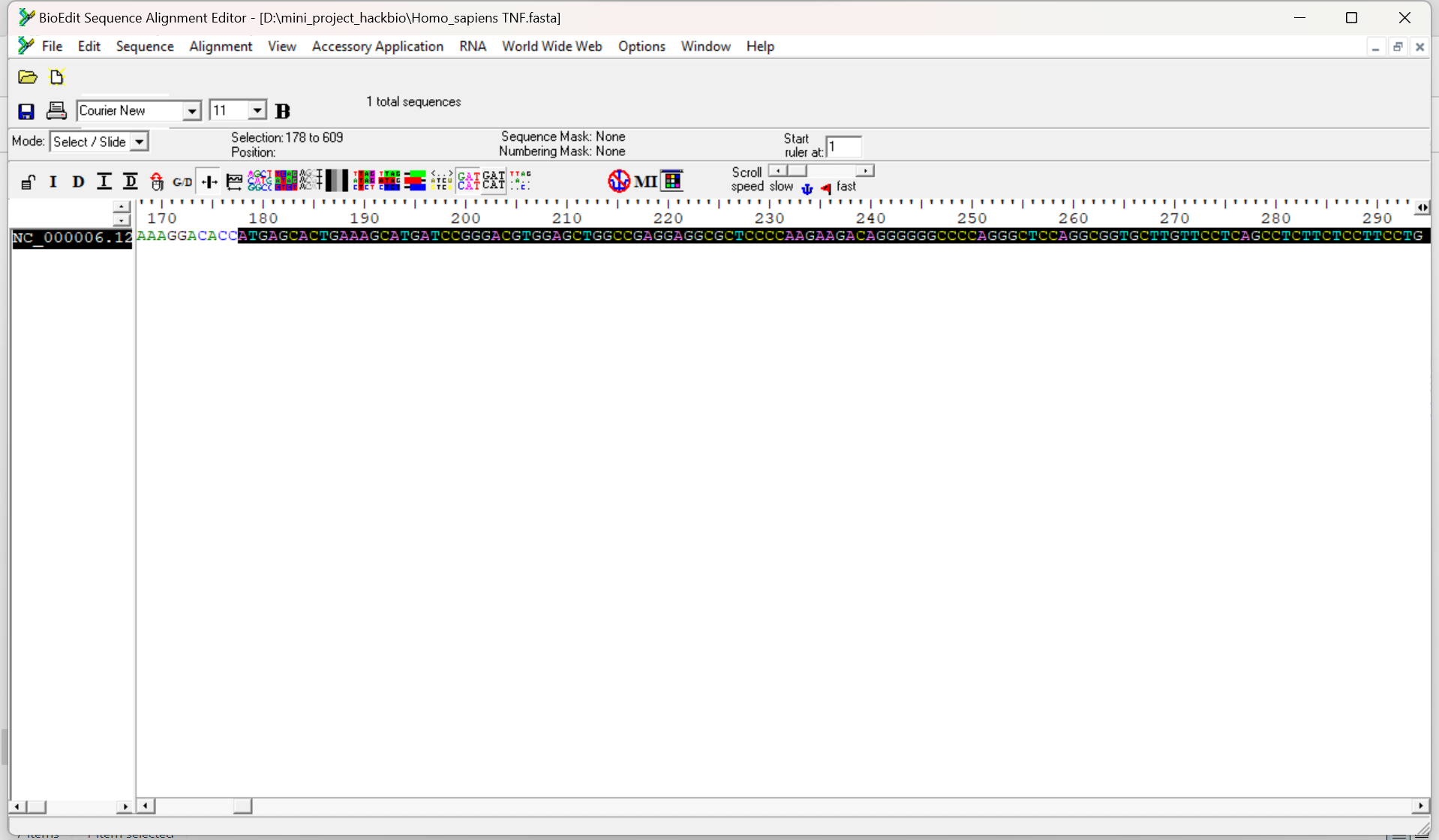


Figure 3

* In Figure 4 the stop codon (TGA) is present.

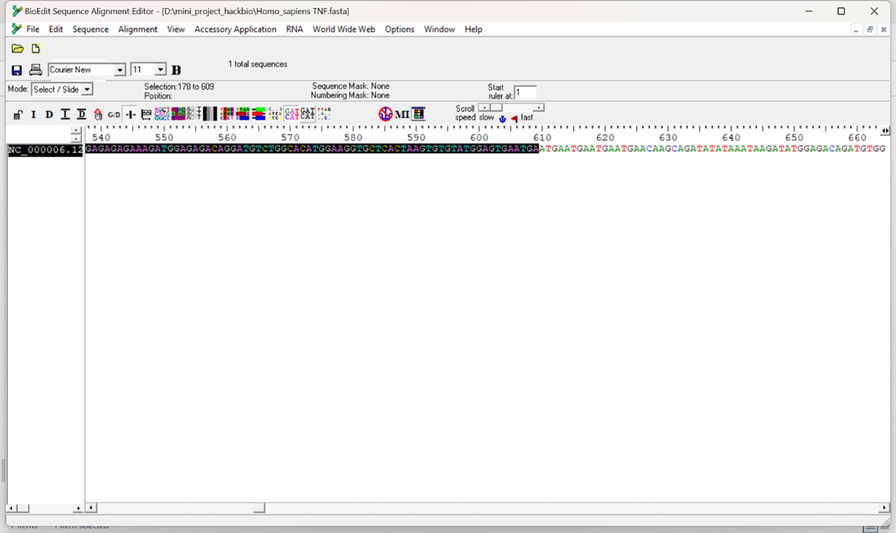


Figure 4

* In Figure 5 we can conclude
* The total length of TNF gene sequence is 924 bases, followed by that is the protein translation.

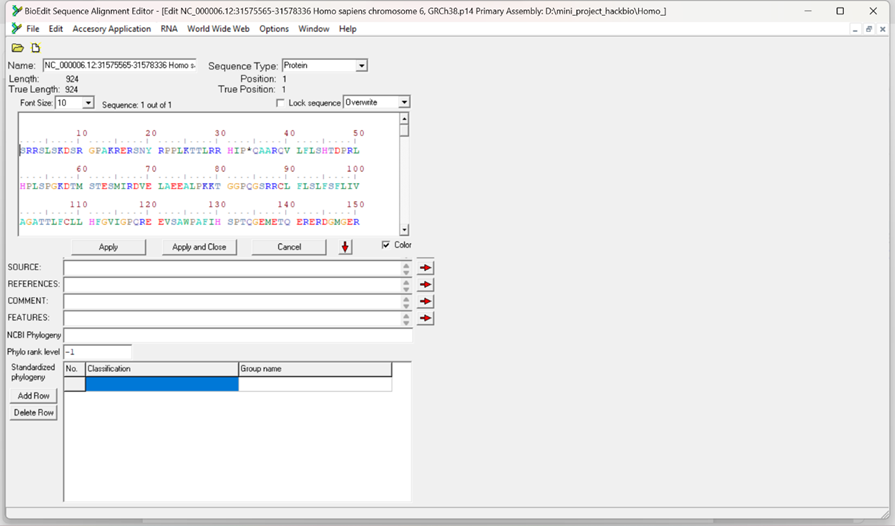


Figure 5

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

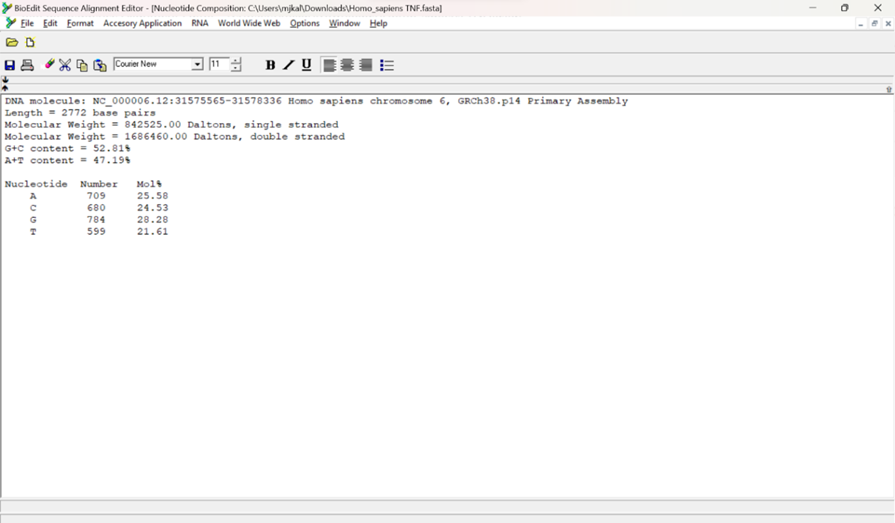


Figure 6

* In Figure 6 we can conclude-

1. The analyzed DNA sequence is 2772 base pairs in length.
2. It has a molecular weight of 842,525 Daltons when single-stranded and 1,686,460 Daltons when double-stranded.
3. The base composition reveals that the sequence contains 52.81% guanine-cytosine (GC) content and 47.19% adenine-thymine (AT) content.
4. The sequence consists of 709 adenine (A) bases, making up 25.58% of the total; 680 cytosine (C) bases, accounting for 24.53%; 784 guanine (G) bases, comprising 28.28%; and 599 thymine (T) bases, representing 21.61%.
5. The sequence has a balanced distribution of nucleotides with a slight bias towards GC content, suggesting potential stability due to the stronger triple hydrogen bonds between guanine and cytosine.

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

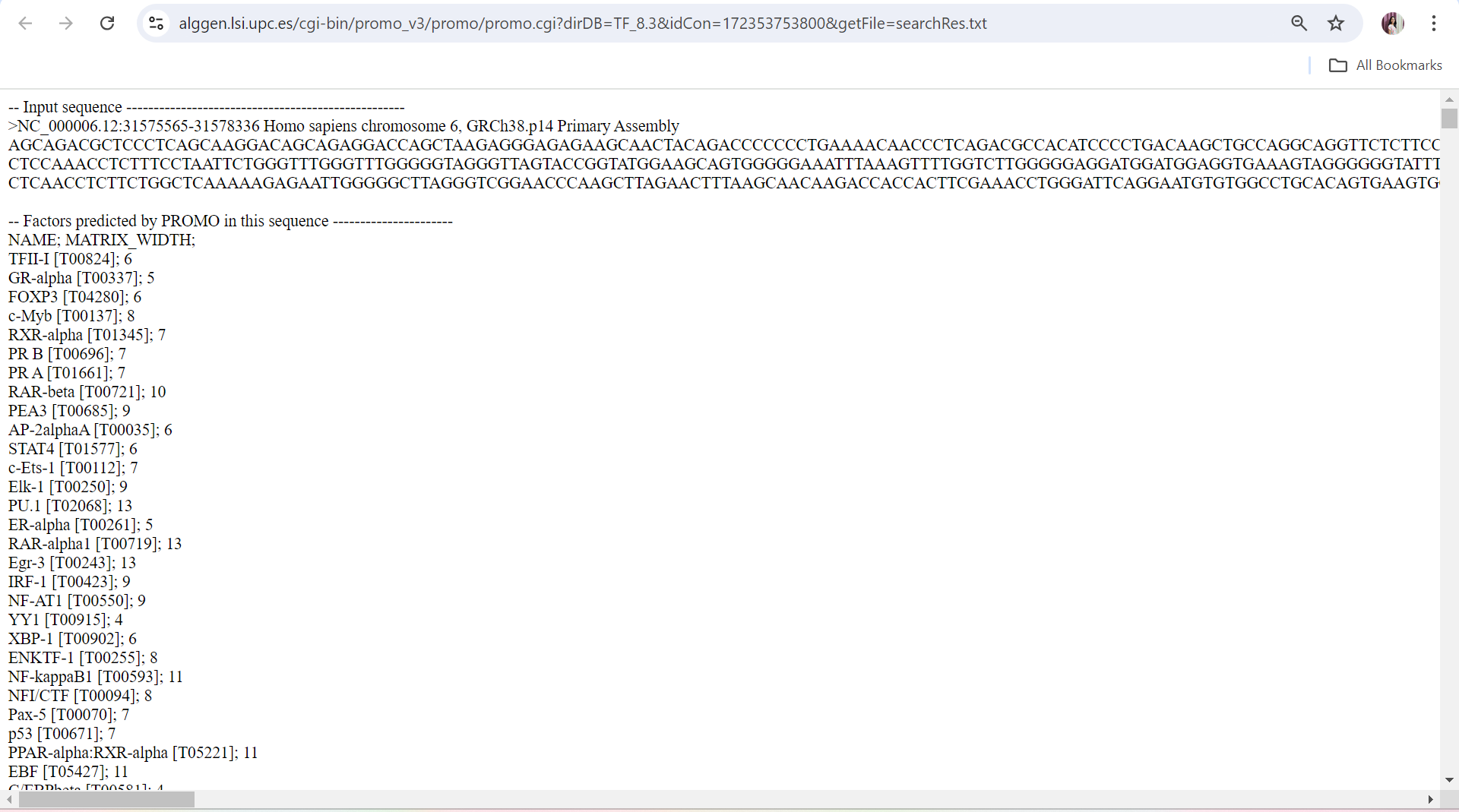


Figure 7

Explanation:

1. The PROMO tool identified several transcription factor binding sites within the input DNA sequence from chromosome 6.
2. Transcription factors include TFII-I, GR-alpha, FoxP3, c-Myb, and RXR-alpha, among others, each with specific binding sites.
3. The matrix width next to each factor indicates the length of the sequence they recognize.
4. These factors play roles in various cellular processes, such as gene expression regulation, immune response, and cell differentiation.
5. The presence of multiple binding sites suggests that this DNA region could be important for complex regulatory functions.

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

Explanation:

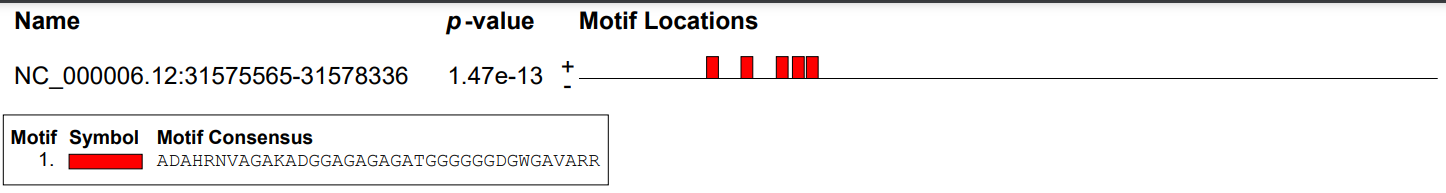


Figure 8

1. Figure 8 is indicating the functional motif region predicted by MEME tool from the TNF gene sequence.

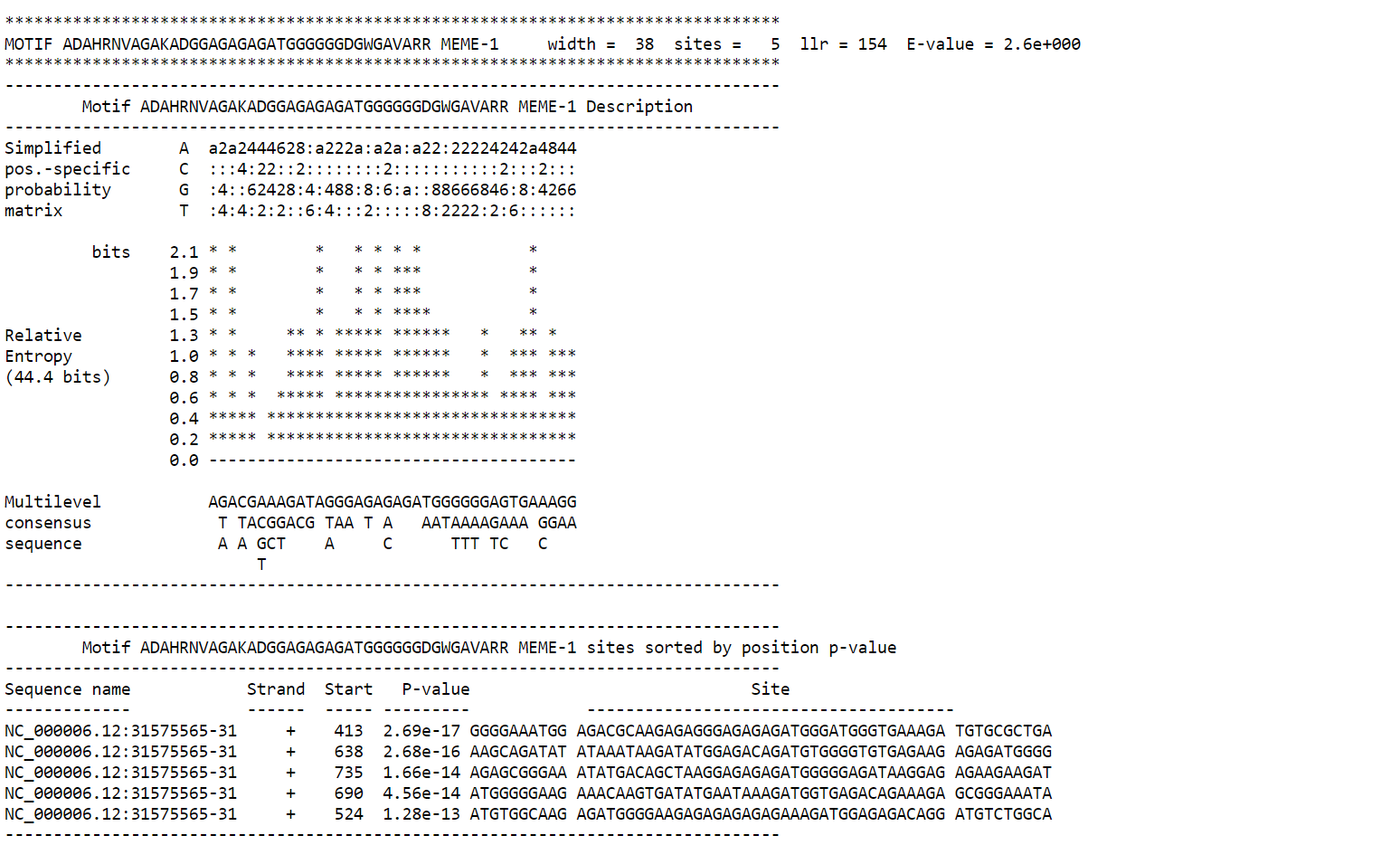


Figure 9

1. In Figure 9 we can see detail explanation including sequence containing the motif, strand (forward or reverse) on which the motif was found.
2. Based on the output, the MEME tool has identified a motif of length 38 nucleotides (or amino acids, depending on your input) that appears in multiple sequences.
3. E-value of 2.6e+000e suggests that this motif is statistically significant and likely represents a biologically relevant pattern.

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

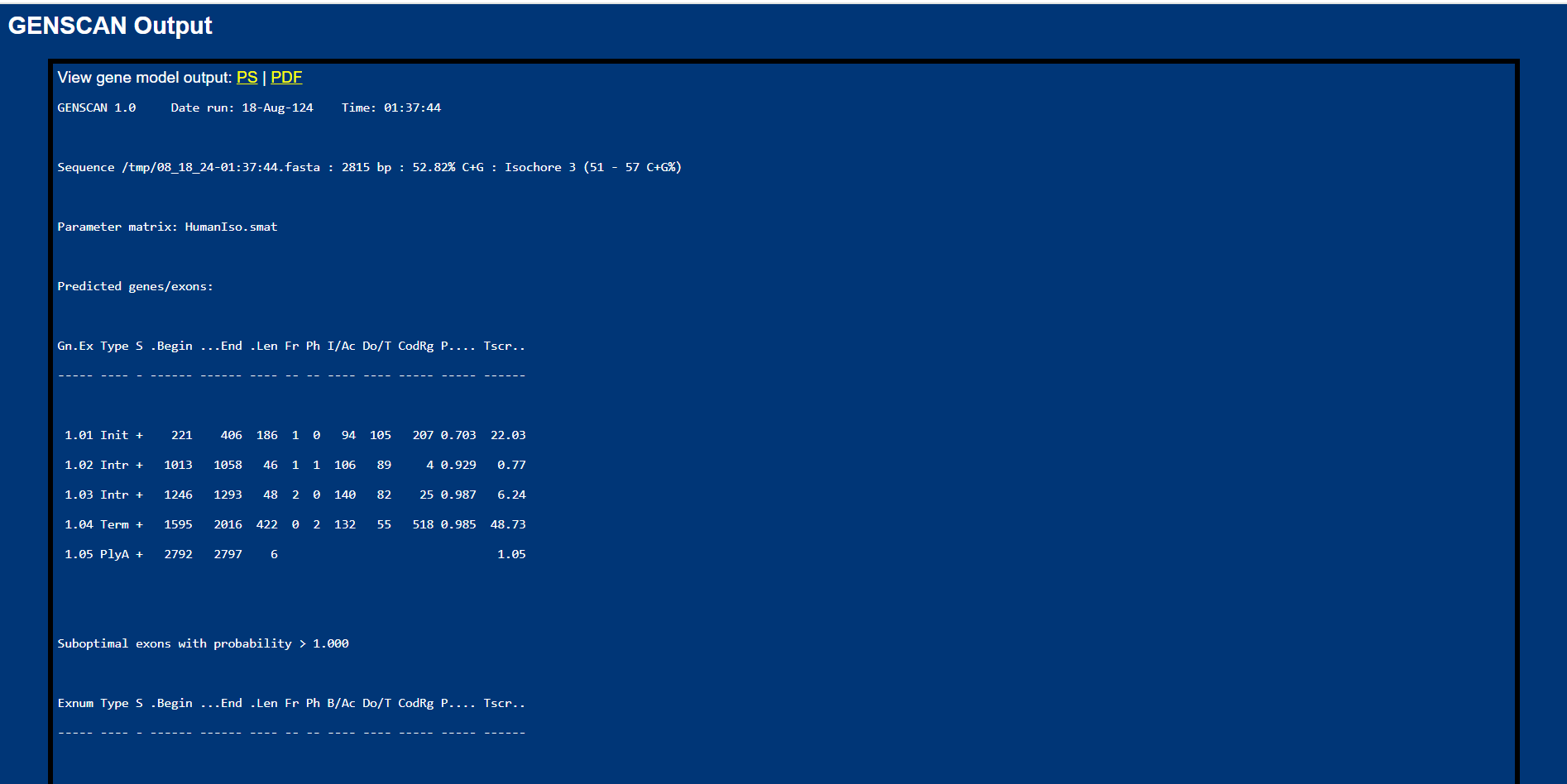


Figure 10

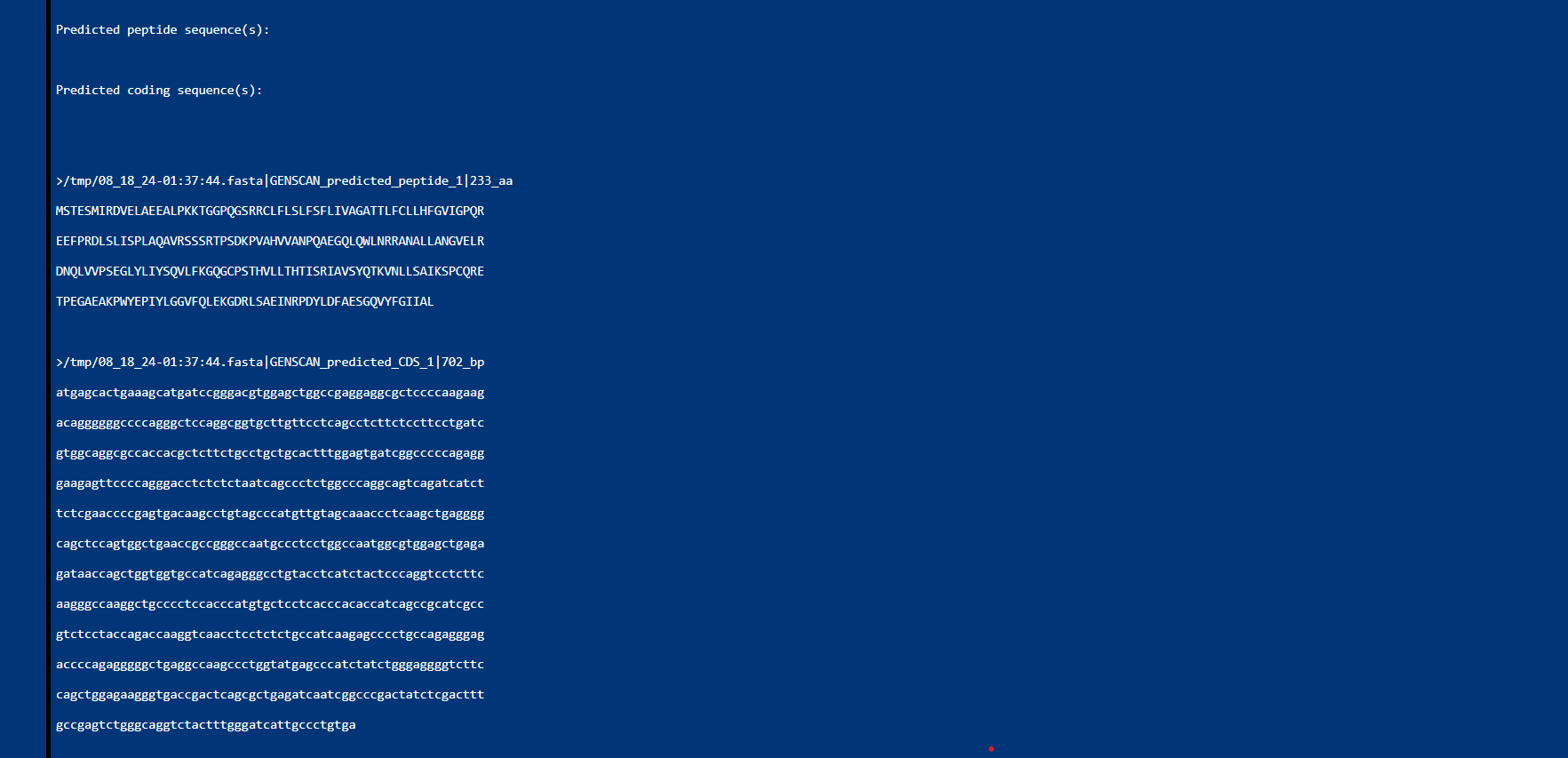


Figure 11

Explanation:

1. Figure 10 is explaining basic information about the input sequence including input sequence name and length and GC content
2. Figure 11 is explaining no exons found at given probability cutoff while it has depicted the predicted coding sequence.
3. A higher cutoff increases specificity, reducing the risk of false positives, but may miss true exons, leading to underprediction.

Hence, we try to search at lower cutoff and we get the results-

1. These exons are less confidently predicted. They may represent alternative splicing events, minor exons, or even spurious predictions.

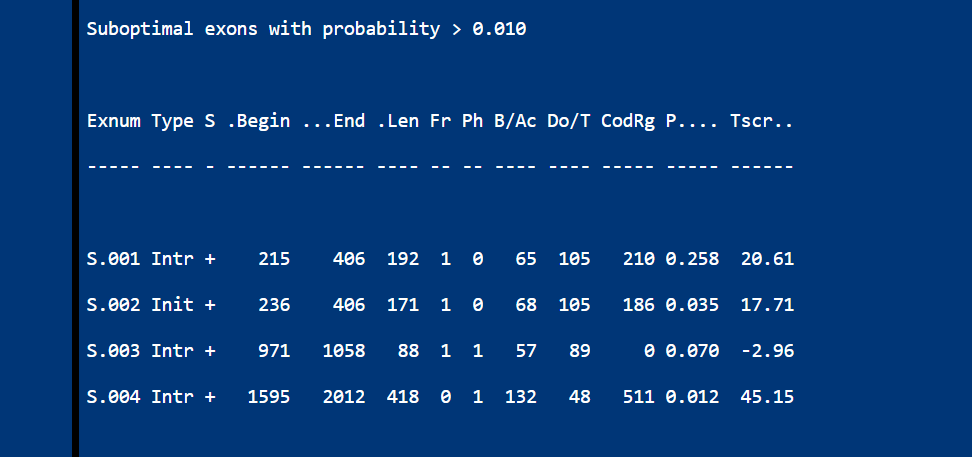


Figure 12

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

1. In Figure 10 we can see the PHYLIP format converted from fasta format.

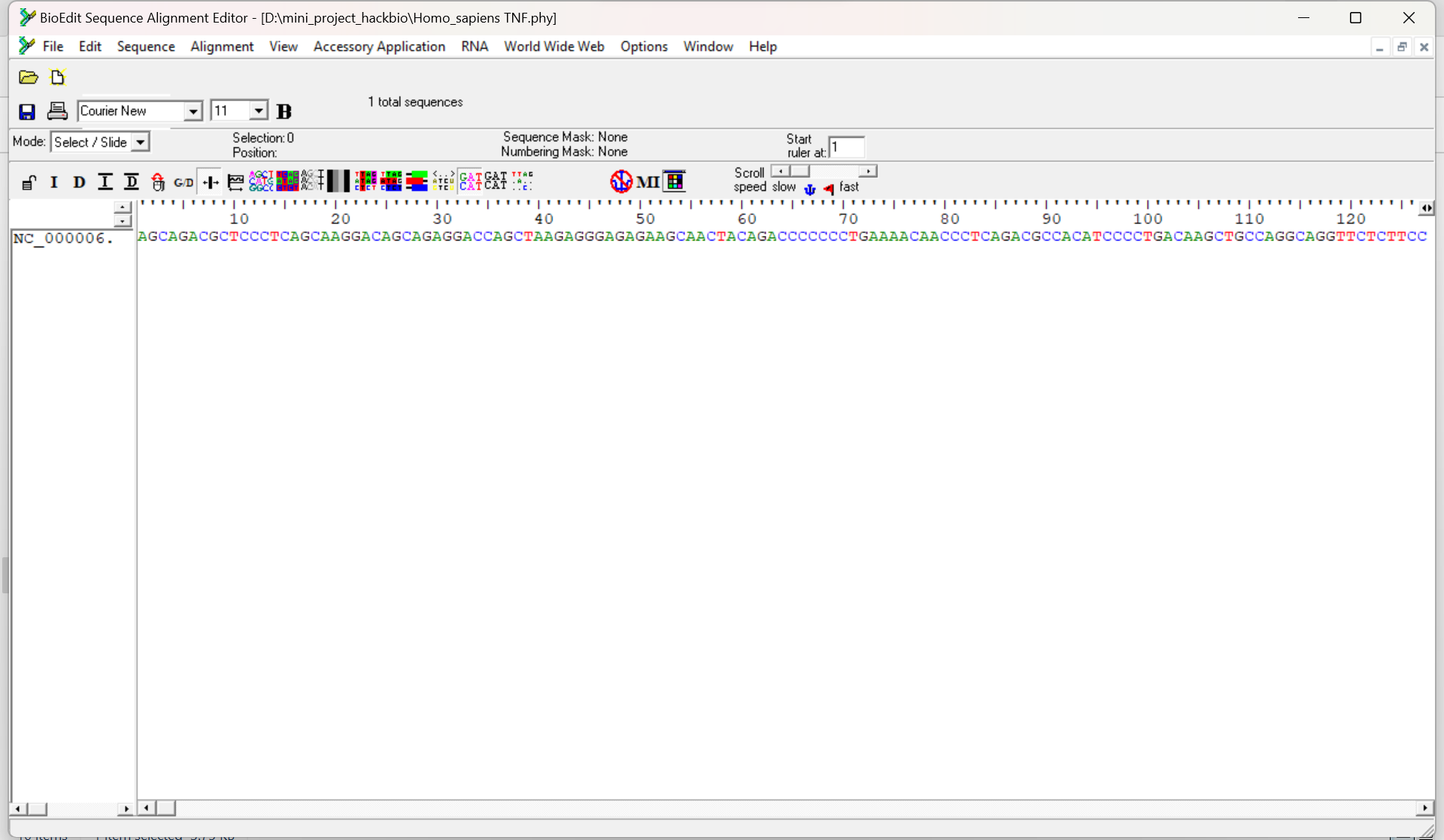


Figure 13