Nicky Lin

Prof. Stetler

BIOL-1015-03

01/25/23

Lab 3

Chart

Description automatically generated

Figure 1. is a segment graph of DNA sequences from the cat sample file (upstream5000cat.fa). The graph shows the range of CPG islands, in blue, for each sequence.

Takeaway

Many of the CPG islands found in this sample of 20 sequences were found towards the end of the sequences. However, one sequence did start in the beginning.

Chart, bar chart, box and whisker chart

Description automatically generated

Figure 2. is a box plot of the GC\_content percentage from 20 sequences from the cat sample file. The box plot shows the distribution of GC\_content percentage.

Takeaway

The bulk of the samples have a GC\_content percentage between 65% and 73%. There does not seem to be any outliers.

Chart, box and whisker chart

Description automatically generated

Figure 3. is a box plot of the lengths of the CPG islands, in nucleotides, found in the 20 sequences from the sample cat file. The box plot shows the distribution of CPG island lengths.

Takeaway

Most of the lengths of the CPG islands fall between ~300 and 650 nucleotides. However, there is an outlier at with a length of 2025 nucleotides.

Sampling Procedure

I collected the 20 samples of cat CPG data by going through each sequence in the upstream5000cat.fa file. If I found a unique sequence, I added it to my student\_data\_sheet.csv file. If no CPG islands were found, I went on to the next sequence in the file.

Tables and Summary

|  |  |
| --- | --- |
| length | (nucleotides) |
| average | 577.0 |
| median | 428.0 |
| minimum | 244.0 |
| maximum | 2025.0 |

|  |  |
| --- | --- |
| GC\_content | (%) |
| average | 68.97 |
| median | 70.45 |
| minimum | 54.90 |
| maximum | 82.70 |

