

Mini Project #1: Where are the Genes is due

Matja Mav (63130148)

November 30, 2019

1 Introduction

In this assignment we had implement the gene-finding algorithm and then use it to analyze the genome of the *Mycoplasma genitalium*.

2 Implementation

We analyzed all open reading frames, using slightly modified translation table 4. We used only one start codon *ATG* and two stop codons *TAA* and *TAG*, since using full translation table resulted into much higher recall.

3 Results

1. **What is the size of *Mycoplasma genitalium* genome?**

580076 bp

2. **How many genes does it include?**

509

3. **What is the length of the smallest and the longest gene (in codons)? What is the median length of the gene (in codons)?**

Shortest length: 38

Longest length: 1806

Median length: 288

4. **What is the recall/precision of your gene finding procedure at L=50 and L=125 codons?**

At length of 50 codons the precision is 0.12 and the recall 0.85.

At length of 125 codons the precision is 0.20 and the recall 0.75.

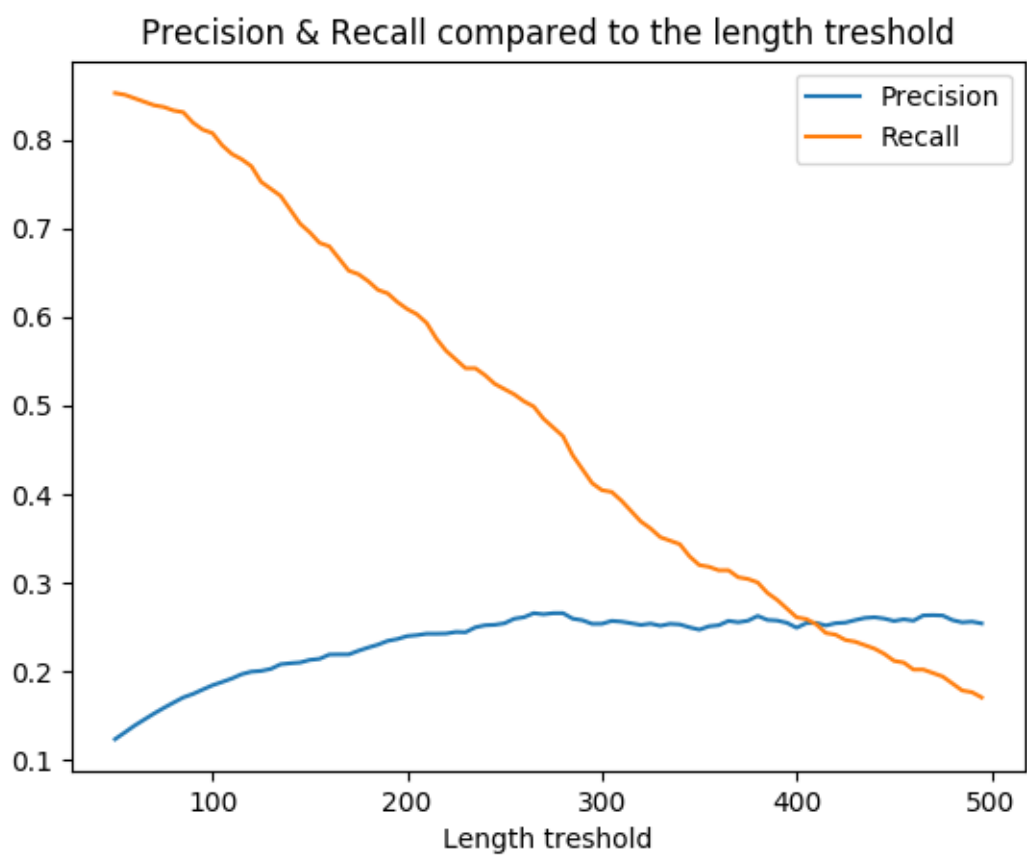


Figure 1: Comparison of the precision and recall to the length threshold