

Analysis report for the "gag" gene in 6 supposedly alike sequences

Luis Jimenez

December 23, 2024

1 Introduction

The following analysis is made by calculating the guanine-cytosine percentage of a specific gene in several viral sequences and comparing them to hopefully observe any similarities. The Spearman test for correlations was also made between some sequences. It is known that viruses are classified in accordance to their characteristics , such as capsid type, DNA chain type, infective capacity, etc. In this case, the gag gene in retrovirus is key in the replication cycle, starting at the formation of new virions until its release and maturation. It codifies for the structural proteins needed for assembly and keeping the integrity of the viral particle. The analyzed sequences extracted from GeneBank are the following: AB097871.1, AB052867.1, KY574515.1, LC735413.1, LC312715.1, DQ093792.1.

2 Results

2.1 Histogram of % GC Content

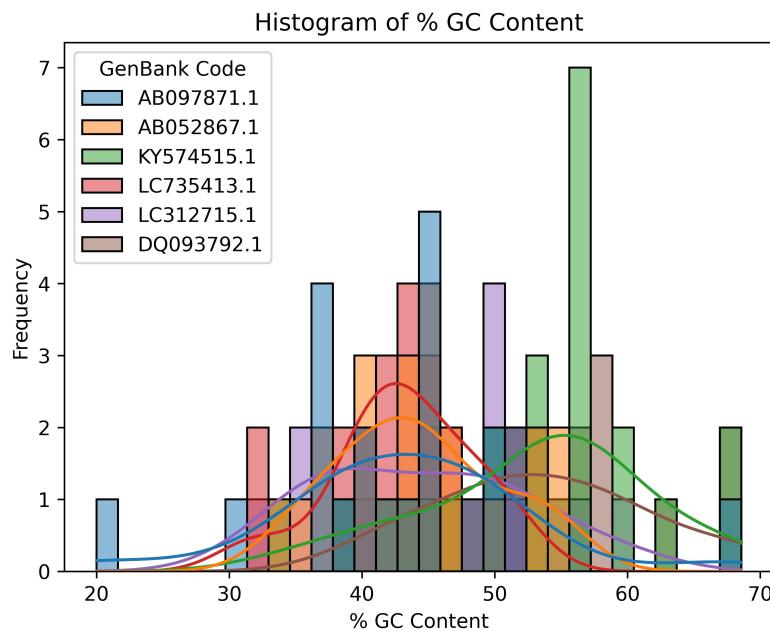


Figure 1: Distribution of % GC content in the sequences analyzed by fragments at a time.

2.2 Density of % GC Content and % GC Content for the randomized sequences

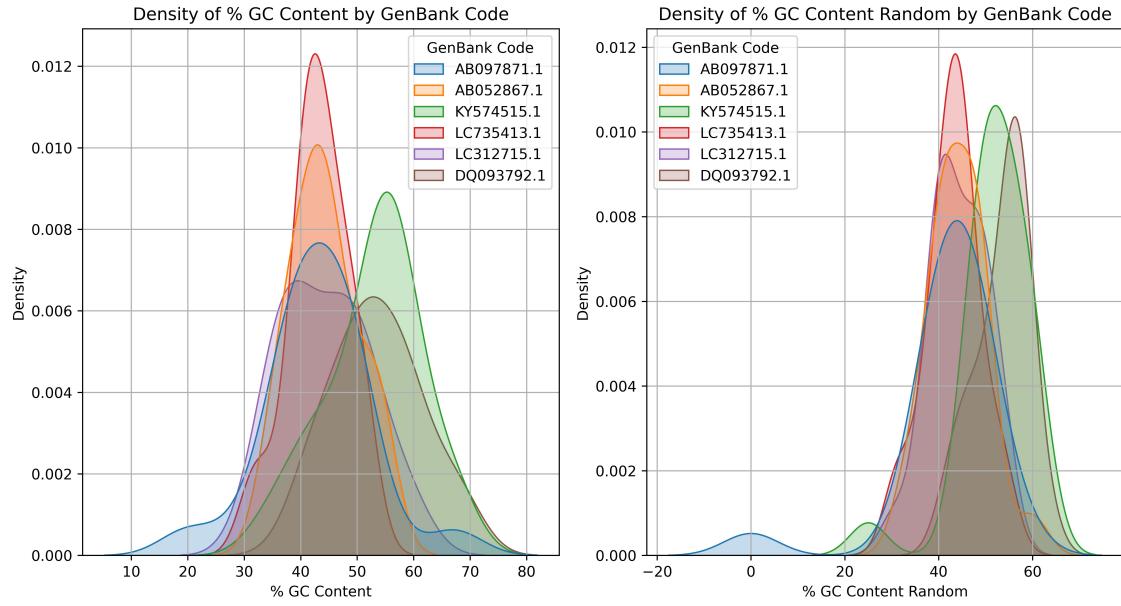


Figure 2: Density of % GC content in the sequences analyzed by fragments at a time.

2.3 Lineplots of % GC Content by position in the sequence

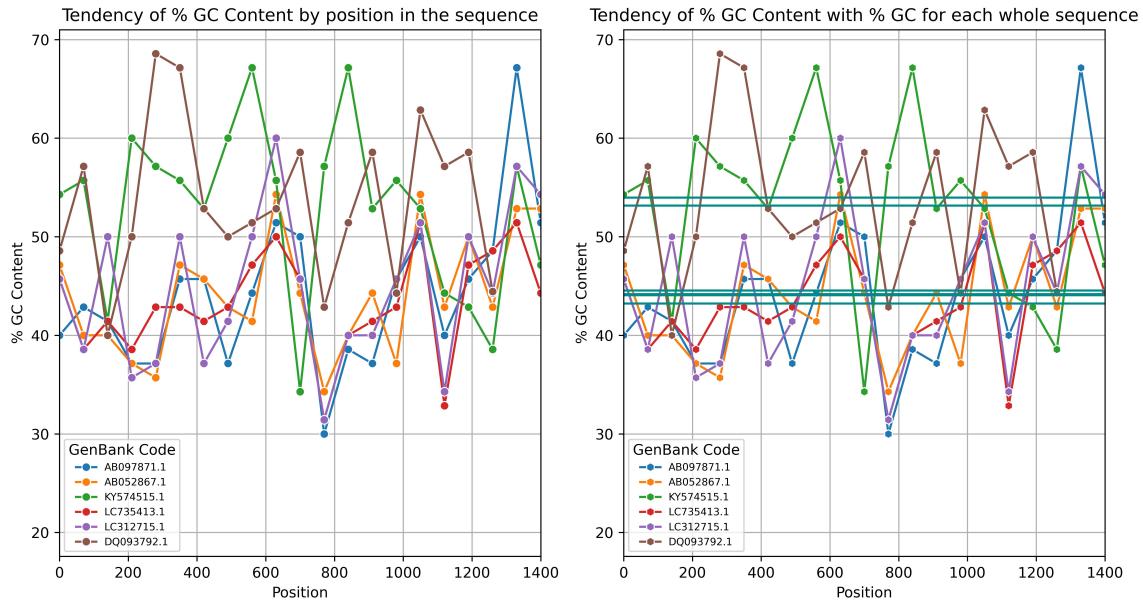


Figure 3: Distribution of % GC content in the sequences analyzed by fragments at a time.

2.4 Violinplots for the distribution of % GC Content by sequence and quartiles

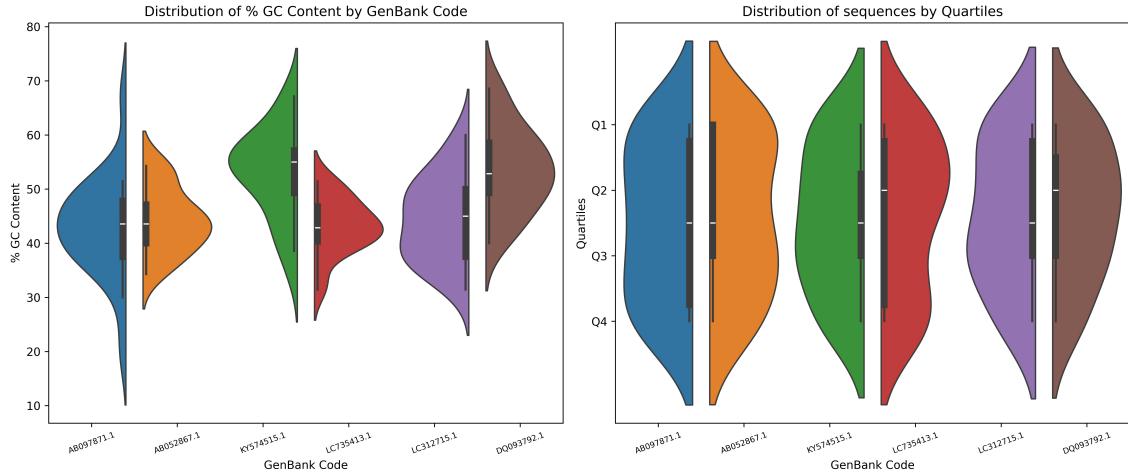


Figure 4: Distribution of % GC content by GenBank Code and quartiles.

2.5 Scatterplot of % GC Content by position and GenBank Code with regression lines

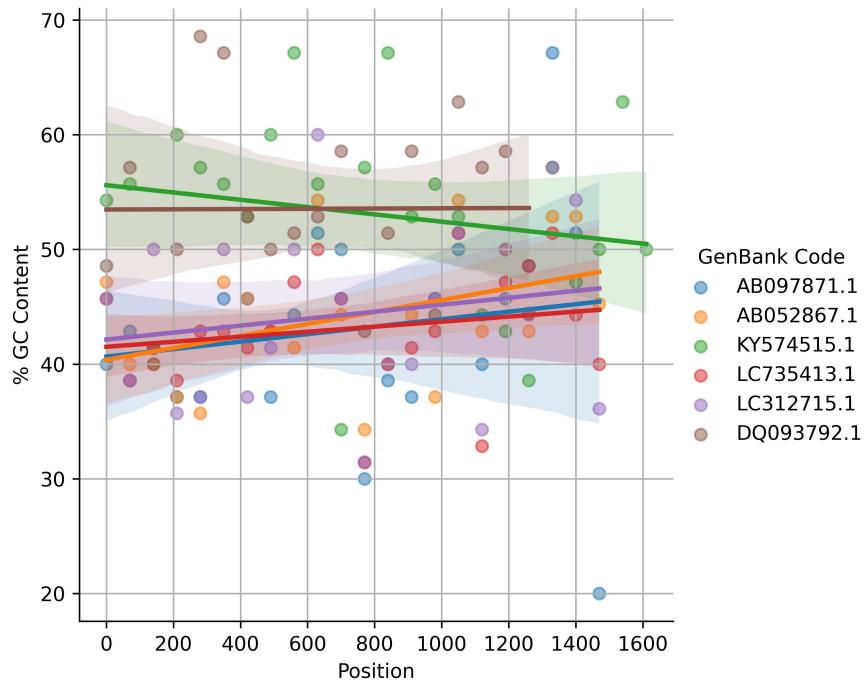


Figure 5: Dispersion and regression lines for % GC Content by position

2.6 Boxplots for % GC Content by quartiles

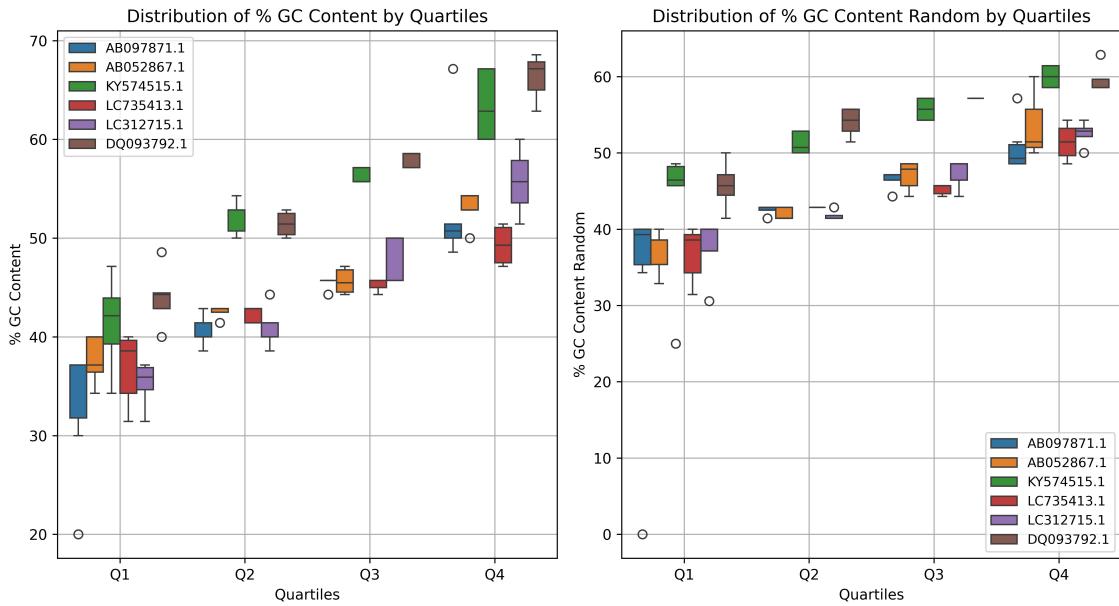


Figure 6: Distribution of % GC Content and Random GC Content by Quartiles

2.7 Heatmap for % GC Content shows similarities

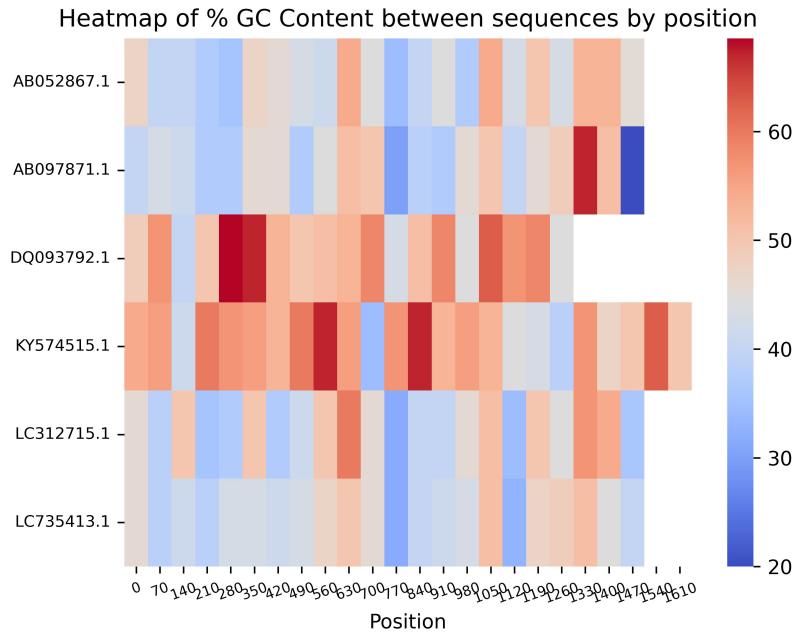


Figure 7: Comparison of % GC Content and GC Content Random

2.8 Bar and line plots for % GC Content and Random GC Content for whole sequences

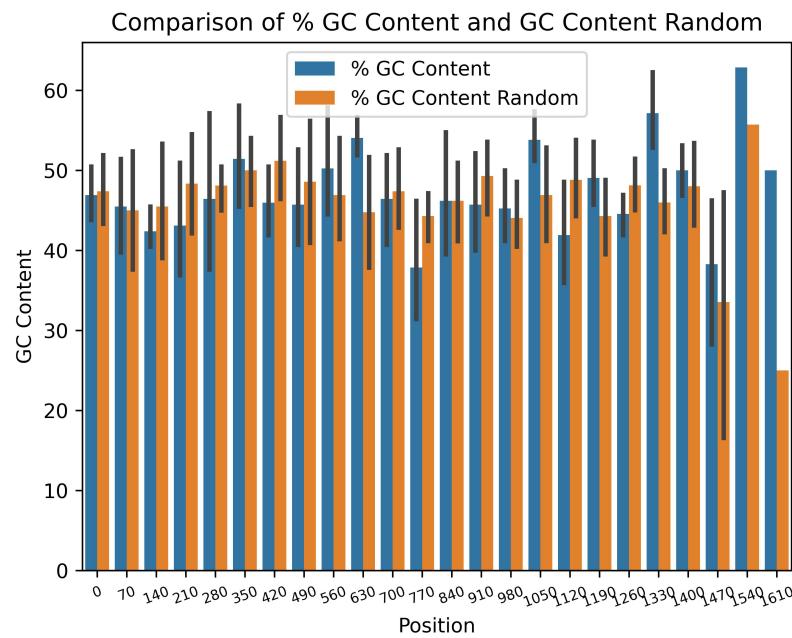


Figure 8: Comparison of % GC Content and GC Content Random - Barplot

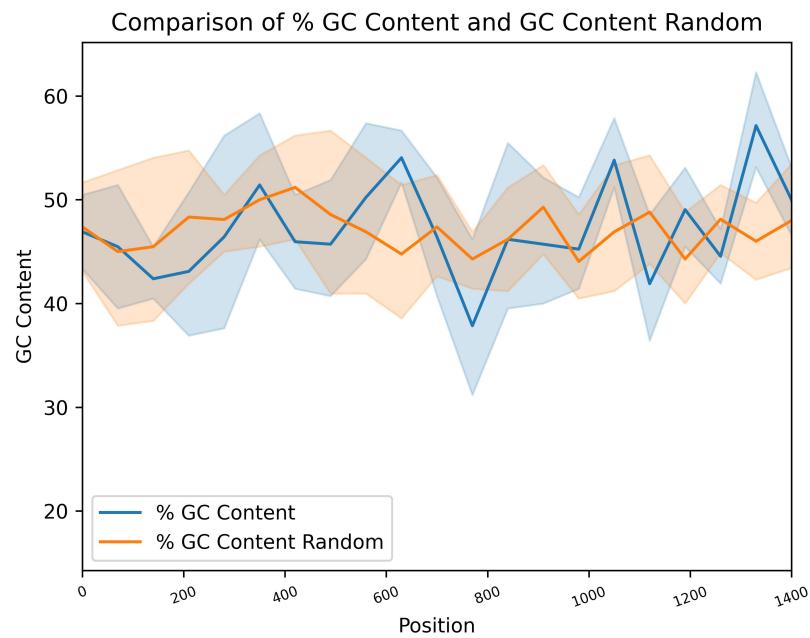


Figure 9: Comparison of % GC Content and GC Content Random - Lineplot