

# GC Base Counts and Their Effects On Genomes

Ignas Ramanauskas

December 14, 2023

## 1 Introduction

This report outlines the methods and findings of a project conducted as part of our GNU/Linux Type Operating Systems project, focusing on the analysis of GC content in yeast genomes. The GC content, representing the proportion of guanine and cytosine bases in DNA, is an important genetic parameter that can provide insights into genome organization and function.

Our approach involved utilizing Bash scripting, a tool used in data manipulation and analysis, to handle and process genome data. We specifically targeted yeast genomes for our study, given their significance in both biological research and industry. The main objective was to examine the variations in GC content across different yeast species, provide acquired data via boxplots and find insights and use this gathered data to compare with scientific literature.

According to A. E. Vinogradov's published paper "DNA helix: the importance of being GC-rich", it was assumed that higher counts of GC correlated with higher DNA thermostability and bendability, thus different ectothermic and exothermic organisms [1].

## 2 Methods

In this project multiple scripts like fasta-unfold, find-orfs and gc-count, as well as python script, were used to analyze and modify data from genome fasta files and provide images with useful data.

In A. E. Vinogradov's paper the genome data was collected from GeneBank. The bendability and curvature of nucleotide sequences were determined using the trinucleotide table of consensus values obtained from DNase I digestion and nucleosome positioning studies. The thermostability was determined using a unified dinucleotide table for free energy of melting.

### 3 Results

According to A. E. Vinogradov's research paper, it was found that higher GC counts did not correlate with higher thermostability, in fact higher counts of GC had lower levels of thermostability.

With increased elevation of GC percentages it was seen that bendability also increased, which would suggest that places in genome with higher counts of GC would result in increased expression.

In the Figure 1. you can see the average percentages of CLIB215 yeast species. We can clearly see that the percentages of CG bases are really similar, around 38 percent. The key difference is that the second image of the GC count have much more outliers than the first one.

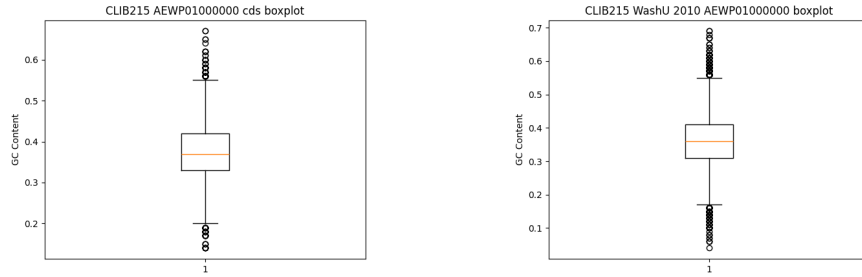


Figure 1: GC counts of CLIB215 yeast species (percentage)

Figure 2. provides us with a similar result as the first one, with slightly higher percentages, around 39 percent in the first image and 35 percent in the second one. Second one also having more outliers.

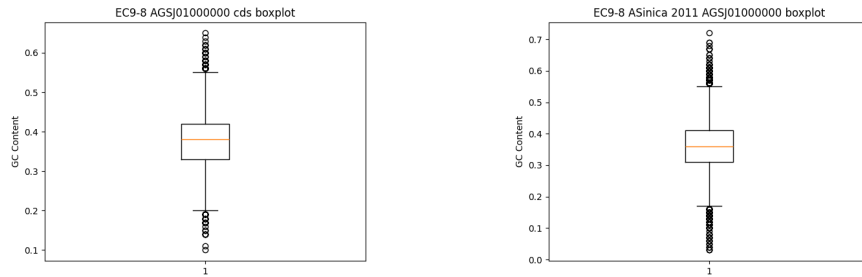


Figure 2: GC counts of EC9-8 yeast species (percentage)

## 4 Conclusions

To summarise our results, according to A. E. Vinogradov, we saw that increased GC percentages in genomes resulted in slightly less thermostability, but with increased DNA's bendability, which could hint us that places with higher GC percentages are more likely to be coding regions.

When comparing GC counts in different yeast species from different databases, we saw a similarity of GC percentages across all yeast species, which was expected. The differences in outliers between same species of yeast could be explained by having different data from different databases and different year that the genome files were collected. For better results, in the future it would be more preferred to use genome files from a similar timeframe.

## References

- [1] Alexander E. Vinogradov. Dna helix: the importance of being gc-rich. *Nucleic Acids Research*, 31(7):1838–1844, 2003. Received January 21, 2003; Revised and Accepted February 12, 2003.