

GC Content in Yeast

Božena Nevinskienė

November 17, 2019

1 About the study

Saccharomyces cerevisiae is a single-celled eukaryote that is frequently used in scientific research. It is an attractive and very popular model organism due to the fact that its genetics are easily manipulated, and it is very easy to maintain in the lab. This single-celled organism is also important in industry, where it is used to make bread, beer, wine, enzymes, and pharmaceuticals. The *Saccharomyces cerevisiae* genome is approximately 12 Mb, organized in 16 chromosomes.

According to the genome databases, GC content of *Saccharomyces cerevisiae* is 38.3%. The goal of this study is to check if it is true. Also to determine if all yeast genes have the same GC content, and if not, why? GC content is very important in experiments, such as PCR, where GC content is important for primer designing, predicting their annealing temperature. Also, GC content showed a quadratic relationship with genome size, with the decreases in GC content in larger genomes possibly being a consequence of the higher biochemical costs of GC base synthesis.

For this study we randomly selected 3 strains of yeast and found their possible open reading frames:

1. EC9-8: found 8622 possible open reading frames.
2. VL3: found 9859 possible open reading frames.
3. YS9: found 8341 possible open reading frames.

In this study, we did not search for reverse complement open reading frames on the antisense DNA chain. Also, not all reading frames may be accurate. The only criteria, that we followed is that the open reading frame must be at least 50 codons in length between start codon (ATG) and stop codons (TAA, TAG and TGA). After finding these ORFs using shell script's "grep", we calculated GC ratio in each of the ORFs (using shell script, sed and awk) and analysed frequency of GC ratios (by drawing histograms in R script).

1.1 EC9-8 results

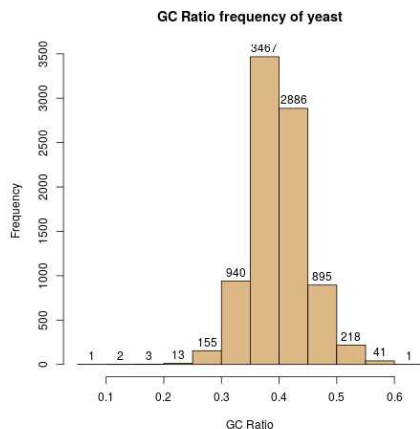


Figure 1: GC ratios of EC9-8

1.2 VL3 results

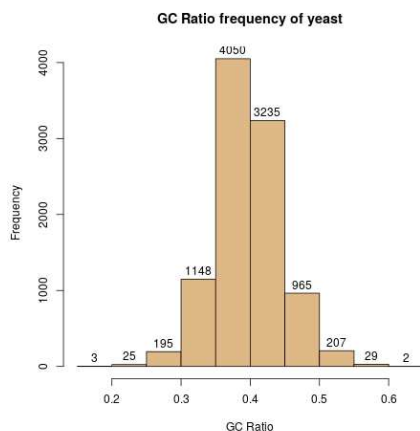
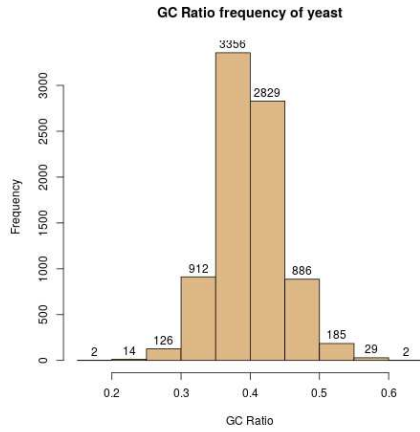


Figure 2: GC ratios of VL3

EC9-8 is a haploid cadmium-resistant derivative of a yeast isolated from the valley bottom of Evolution Canyon at Lower Nahal Oren, Israel. This yeast strain had the highest number of outliers in GC ratio. But on the average of G and C nucleotides in EC9-8 yeast strain's possible ORFS was approximately 39.94%. More than 70% of ORFS had the G and C ratio between 35% and 45% (Fig. 1).

VL3 was isolated in Bordeaux, France, and is most suited to the production of premium aromatic white wines with high thiol content (citrus and tropical fruit characters). VL3 has a whole-chromosome amplification of chromosome VIII, as well as 54 ORFs that are missing in other strains. The average of G and C nucleotides in VL3 yeast strain's possible ORFS was approximately 39.67% (Fig. 2).

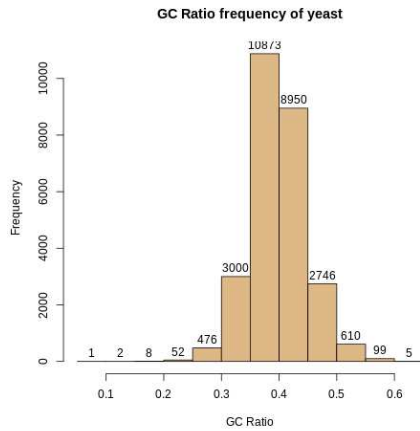
1.3 YS9 results



YS9 is one of the Singapore's baking yeast strains. The average of G and C nucleotides' ratio in EC9-8 yeast strain's possible ORFS was approximately 39.97% (Fig. 3).

Figure 3: GC ratios of YS9

2 Conclusions



Overall in all 3 strains the mean of G and C nucleotides was about 39.86% of all nucleotides (Fig. 4). According to Anderson-Darling normality test, none of the GC ratio results were distributed normally. Although, we have a large sample size, so we can compare GC ratios between 3 strains with a T test. Results have shown, that GC ratios of strains YS9 and EC9-8 are not significantly different. But both of those strains' GC ratio means significantly differ from VL3 strain's.

Figure 4: GC ratios of all three strains

All in all, the most possible open reading frames had GC ratios between 0.35 and 0.45. The GC-content of all three strains was similar, but not perfectly 38.3% as it was predicted. This 1.56% difference seems very small but it is significant. In the future studies we should consider including reverse complement open reading frames for more accurate results.