

Question: Analysis of tRNA genes found by only Aragorn.

Alignment of genes found by Aragorn only:

> Before aligning the Aragorn genes, I removed the introns and variable **arms**.

For sequences with secondary structure like this: ((((((.(((ddddd))) .(((AAA))))vvvvv..... (((((((ttttt))))))))))

which had no arm, I did not remove the variable part because I was not sure which positions are 45-48.

I assumed that the variable part will be removed in the alignment with the new criteria defined in the alignment pipeline.

> 741 genes out of 4321 genes were found only by Aragorn. These were aligned together with the genes in the intersection set using coeva.

> In editing the alignment, instead of removing sites with more than 99% gap, I removed sites with more than 83 gaps $(4321-741)/4321$. also, genes with more than two gaps in a row and more than 3 gaps in total were removed.

The result did not change the number of genes and the consensus secondary structure of genes in intersection set (I compared the result with the result of the alignment of intersection sets only).

Although from 741 Aragorn genes there were only 447 genes left.

> Positive and negative profiles were created using the odd ratios. I calculated the score of each gene against positive and negative profile. Figure 1 shows the distribution of scores against positive model (colored in blue) and negative model (colored in red).

> Then I plotted the distribution of scores for 447 genes found by Aragorn against positive and negative profile. Figure 2 shows the distribution of foreground scores (scores against positive model) with color yellow. Figure 3 shows the distribution of scores against negative model with color green.

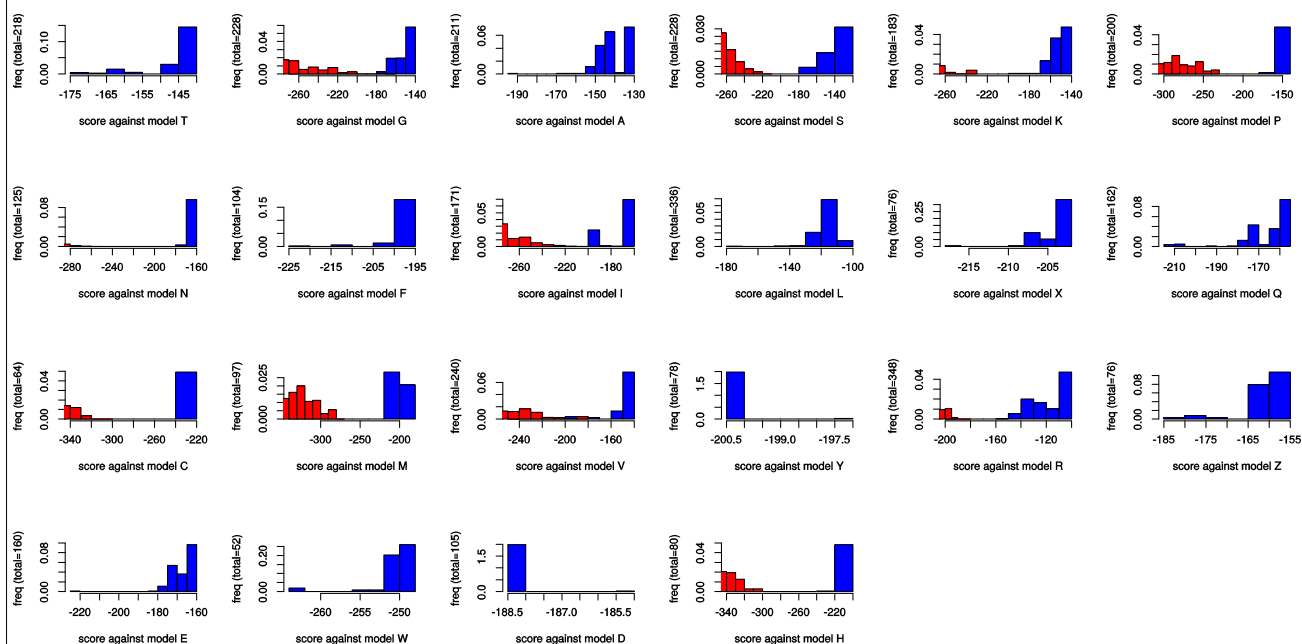


figure 1: Scores of genes in intersection set against positive (blue) and negative (red) profile.

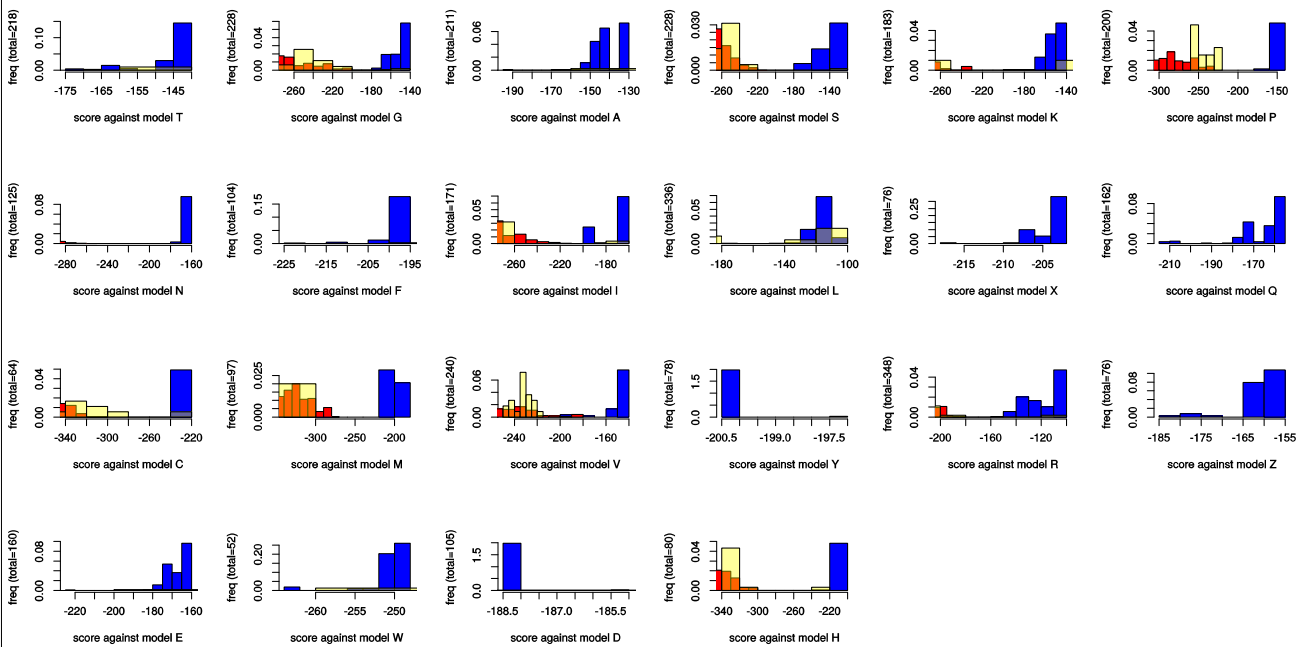


figure 2: Foreground distribution of scores for genes found by Aragorn only (scores against positive profile) colored in yellow added to figure 1. This shows that many of the Aragorn genes may not be true genes.

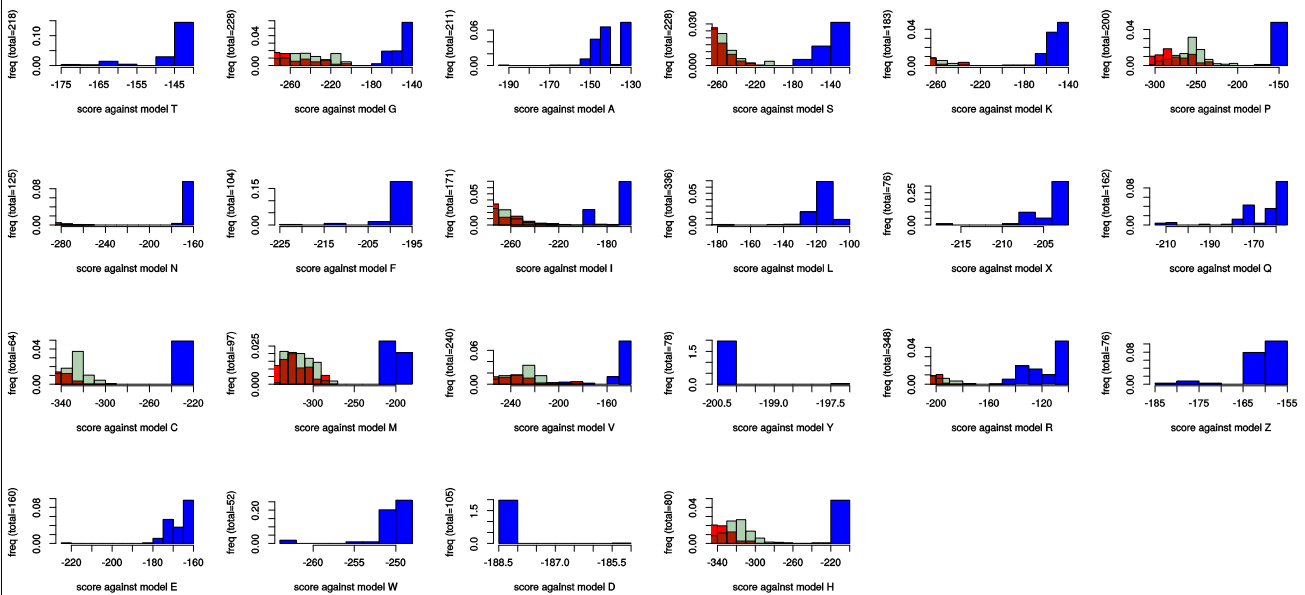


figure 3: Distribution of scores for genes found by Aragorn only against negative profile colored in green added to figure 1.

> Based on these figures, I decided to keep those Aragorn genes whose score against positive profile is within the range of scores colored in blue (foreground set) or higher, and not within the range of scores against negative profile (red color) or lower. Based on this criteria there were only 28 genes left out of 447 genes.