

DnaA box in *Salmonella enterica*

Alejandro Amador

2020/06/09

Following the training in the book, I first made the skew plot (Figure 1) to try finding roughly the *ori* location according to the "decreasing to increasing" behaviour of the plot. In this case, it can be seen that *ori* could be located in the interval $(3.6e6, 4e6)$, since the minimum is located at position 3,764,856.

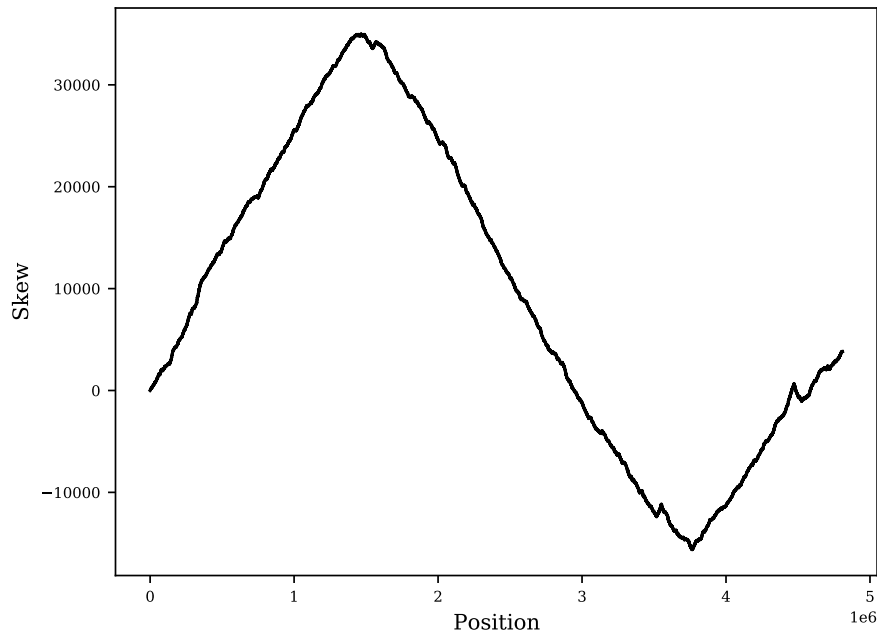


Figure 1: Skew diagram for the *Salmonella enterica*.

After that, I looked for the most frequent 9-mers in a neighborhood of distance 3 and considering reversed complements. The two most frequent were **CCGCCGGCG** and **CGCCGGCGG**. Since the genome is very large, it is not easy to visualize this 9-mers in it by just making their characters bold in the corresponding appearances. Instead, I made a new function to find each appearance and save its position in the genome. With this, we can visualize the clusters as points around the minimum of the skew minimum value (Figure 2). In this plot it can be seen that these two patterns do form clusters around the hypothesized *ori* location, which could indicate that both 9-mers play an important structural rule.

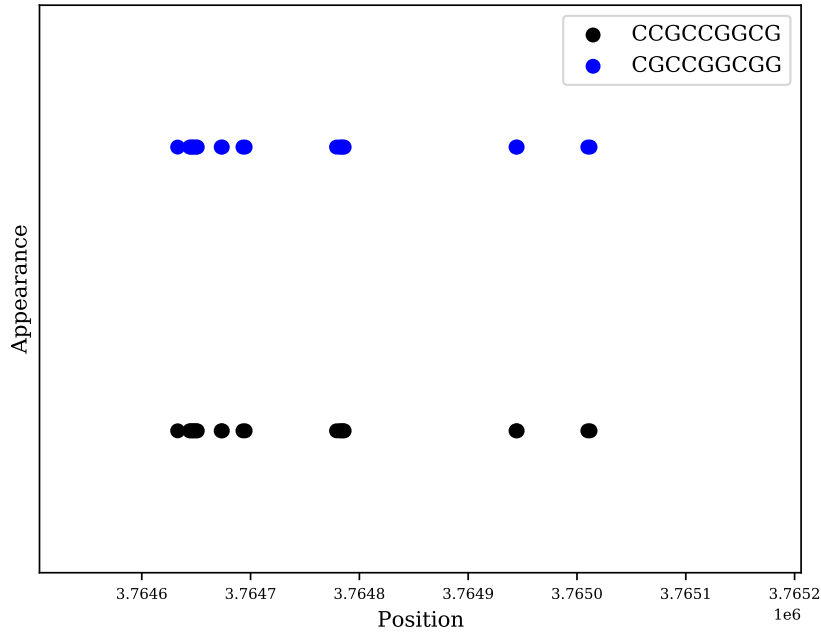


Figure 2: Visualization for the appearance of each 9-mer around the position which minimizes the skew. Each point represents that the corresponding 9-mer appears at that position. They appear at different y-axis position just for visualization.

Moreover, it is noticeable that both 9-mers are very similar to each other, the second one appears to be the last eight characters of the first plus a "G", additionally, they form the same clusters around the minimum. When comparing the appearance positions it is found that they coincide at 4,598,514 positions (their 3-neighbors are the same) and differ at 210,523 positions. This indicates that maybe this two 9-mers are actually part of a longer 10-mer. However, when looking for the most frequent 10-mer, the algorithm returns the patterns **CGCCGGCGGC** and **GCCGCCGGCG**. Each 10-mer contains one of the previously found 9-mers, nonetheless, they show that the 9-mers are not merely different parts of a

larger unique 10-mer and, while the 10-mers are somewhat similar, they do not appear to be part of a larger k -mer ($k \geq 11$), this, however, could be further studied¹.

¹I wanted to do it, but for 10-mers the algorithm was already taking too long and I wanted to progress with the textbook.