



## BP-08: Exploring the Impact of Some Parameters and Their Interactions on the MAFFT MSA Tool Using the Design of Experiments

ROUABAH Safia <sup>1</sup>, DAAS Mohamed Skander <sup>1\*</sup>, BOULAHROUF Khaled <sup>1</sup>, HAMIDECHI Mohamed Abdelhafid <sup>1</sup>

<sup>1</sup> Microbiological Engineering and Applications Laboratory, University of Freres Mentouri Constantine 1, Faculty of Natural and Life Sciences.

*Email\**: [safia.rouabah@doc.umc.edu.dz](mailto:safia.rouabah@doc.umc.edu.dz)

### Abstract

Multiple Sequence Alignment (MSA) is essential in bioinformatics for identifying conserved regions and evolutionary relationships among biological sequences. Due to its efficiency and precision, the MAFFT algorithm is a popular tool for conducting MSA. However, the effect of numerous parameters and their interactions on some performance metrics remains relatively underexplored.

In this study, we investigate the effects and interactions of four important parameters: number of sequences, sequence length, insertion rate, and deletion rate, on four performance metrics of the MAFFT tool.

By generating a diverse dataset of biological sequences, we carried out a comprehensive analysis of MAFFT's performance in terms of Sum of Pairs Score (SPS), Column Score (CS), and Delay. Through a series of controlled experiments using the design of experiments, we assessed the impact of parameters' variation and their interactions on these performance metrics.

Our findings indicate that the considered parameters and their interactions significantly influence the MAFFT's performance across all the metrics. Specifically, the most influential parameter in terms of SPS and CS quality is the number of sequences. However, the sequence length parameter has a greater impact on the delay metric. Additionally, insertion and deletion rates, has a relatively lower impact on all alignment quality metrics.

These results emphasize the importance of parameter impact and their interactions on the MAFFT tool. The study provides insights into the interplay between MAFFT's parameter settings and its performance, enabling researchers and practitioners to make informed decisions when applying the tool to biological sequence alignment tasks.

**Keywords:** Multiple Sequence Alignment, MAFFT, Parameter interaction, Performance metrics, Design of experiments