

# Sequence Alignment: Comparative Analysis of Algorithms in KRAS Genetic Mutations

Ana Rita Pereira<sup>1,2</sup>, Carlos Lopes<sup>1,3</sup>, Catarina Oliveira<sup>1,4</sup>, Gonçalo Pereira<sup>1,5</sup>, Rui Moreira\*<sup>1,6</sup>.

Brígida Mónica Faria<sup>1,7</sup>

1. ESS – Polytechnic of Porto, Portugal, 2. Luz Arrabida Hospital, Vila Nova de Gaia, Portugal; 3. Armed Forces Hospital – Porto Campus, Portugal, 4. Alto Ave Local Health Unit, EPE – Senhora da Oliveira Hospital, Guimarães, Portugal 5. Local Health Unit Santo António – Hospital Magalhães Lemos, Porto, Portugal 6. Local Health Unit São João, EPE – Nossa Senhora da Conceição Hospital, Valongo, Portugal; 7. LIACC – Artificial Intelligence and Computer Science Laboratory (member of LASI), Porto, Portugal

\*Corresponding author email: 10230714@ess.ipp.pt

**Introduction:** For this study a program was developed in Python on biological sequence alignment, considering the application of algorithms in the genetic analysis of Kristen rat sarcoma viral oncogene (KRAS) and its main mutations associated with cancer. The KRAS gene, like other genes in the same family, is responsible for encoding proteins that regulate cell proliferation, differentiation, and apoptosis. The algorithms employed include the Needleman-Wunsch algorithm as well as the Smith-Waterman algorithm and the Basic Local Alignment Search Tool (BLAST). The algorithms for multiple sequence alignment help to understand the function, evolution, and variability of biological sequences, significantly contributing to advances in genomics and proteomics. **Objectives:** The objectives of this study are to apply algorithms for the effective alignment of biological sequences, compare the non-mutated KRAS sequence with principal mutations associated with cancer development, delineate and justify the selection of the algorithms used, assess their computational complexity, and facilitate 3D visualization of the sequences. **Methods:** Development of a program – **BioAlign** – in Python, with various functions including upload and visualization of sequences; use of different algorithms for global and local alignment; BLAST search; algorithms complexity analysis; obtaining the nucleotides positions; obtaining subsequence and its position; phylogenetic analysis; histogram visualization of sequence length and 3D structure visualization. **Results:** The program is capable of analyzing and comparing the provided sequences using both local and global algorithms. The execution time among the three main algorithms differs, with the BLAST algorithm notably slower in returning results. This fact may be due to several factors, such as the complexity of the algorithm itself, the internet speed, and the response time of the NCBI website. **Conclusions:** The development of the **BioAlign** program indeed allows to address the proposed objectives. Furthermore, the completion of this project has enhanced proficiency in utilizing the Python programming language, demonstrating significant skill development.

**Keywords:** sequence alignment; algorithm; KRAS.

## References:

1. Macedo, D. Alinhamento de sequências biológicas na Bioinformática. Centro Universitário Anhanguera de Niterói; UNIAN Niterói; Rio de Janeiro; 2014.
2. Mariano, DCB (org.) et al. BIOINFO – Revista Brasileira de Bioinformática e Biologia Computacional. 1ª Edição. Volume. 1. ISBN: 978-6-599-275326. Lagoa Santa: Alfabelix,; 2021. DOI: 10.51780/978-6-599-275326
3. Alinhamento global (ulisboa.pt) – Consultado em: 08/04/2024
4. National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov> – Consultado em: 09/04/2024
5. Peter M. K. Westcott, Minh D. To. The genetics and biology of KRAS in lung cancer. *Chin J Cancer*. 32(2):63–70; 2013. doi:10.5732/cjc.012.10098
6. Zanatto, R. M., Santos, G., Oliveira, J. C., Pracucho, E. M., Nunes, A. J. F., Lopes-Filho, G. J., & Saad, S. S. Impact of kras mutations in clinical features in colorectal cancer. *ABCD. São Paulo: Arquivos Brasileiros de Cirurgia Digestiva*; 33; e1524; 2020.
7. Lopes, R. M. D. S. C. Marcadores moleculares do adenocarcinoma do pulmão. 2012.