GenTreat: Computational pipeline to perform automated hybrid assembly

Gislenne da Silva Moia, Mônica Silva de Oliveira, Pablo Henrique Caracciolo Gomes de Sá, Jorianne Thyeska Castro Alves, Adonney Allan de Oliveira Veras, Victória Cardoso dos Santos

UNIVERSIDADE FEDERAL DO PARÁ, UNIVERSIDADE FEDERAL DO PARÁ

Abstract

Since the development of the NGS (Next-Generation Sequencing) technologies, the sequencing process of genomes has become faster, with low cost and data generation with high accuracy. To handle a large amount of data and perform different analyzes, for example, assembly, many software have been developed. Among the approaches used to assemble genomes, there is the de novo assembly that performs the task without using a reference organism. Over the years, numerous tools to perform this task were developed, for example, Velvet, Mira, Spades, Abyss, Allpaths. Despite the accuracy of these tools, it is still necessary to perform several assembly rounds, aiming for the best result, it is also possible to use a hybrid strategy where results from different assemblers are used in a joining process to improve the final result. GenTreat is a computational pipeline, with a friendly graphical interface, that performs automated hybrid assembly of prokaryotic genomes, where the assembly result obtained using SPAdes and MegaHit softwares are used as inputs for CISA contigs integrator, the final result is ordered with the Ragoo software and submitted to assessment through of QUAST software, the pipeline accepts as inputs single-end and paired-end reads. To demonstrate the pipeline efficiency, it was executed assemblies using the softwares: Spades, Megahit, Velvet, and the GenTreat, the results showed that the assemblies obtained through of the pipeline are less fragmented, show greater genomic features from analyzing with QUAST, N50 with more value than the others, contigs with lengths greater than 200 base pairs and with the number of bases closest to the expected size of the genome. Therefore, it turns out that the pipeline is a viable alternative to perform hybrid assemblies, in addition to being a tool that accomplishes an automated task without the need for the user to make use of complex and extensive command lines.

Funding:
Link to Video: