Autoassembly: A Tool for the automatic assembly process of prokaryotic genomes

COIMBRA, N. A. R.[1] : LIMA, A. R. J.[1] : de SA, P. H. G.[1] : RAMOS, R. T. J.[1]: SILVA, A. L. C.[1]:

[1]Laboratório de Polimorfismo de DNA – LPDNA, da Universidade Federal do Pará -UFPA, Belém.
{nilson.coimbra, pablogomesdesa}@gmail.com, alexranieri@hotmail.com,{asilva, rommelramos}@ufpa.br

Keywords: genome assembly, pipeline , Java.

Since 2005, after the appearance of the next generation sequencers, platforms able to decode a complete genome in a few days, with high accuracy and coverage sequencing have provided new advances in genomic research. Regarding the exponential increase of the deposited sequences in the NCBI database, the development of software tools, such as MAUVE assembly (Darling, *et al.,* Genome Res, 2004), program that align sequences against the reference genomes and pipelines to assembly genomes are necessary to reduce the processing time of the data produced by these platforms, and also improve the handling of the programs by biologists, for obtaining the genome scaffold. Therefore, this work aims to develop a software tool to automate the process of whole genome assembly, from the quality filter stage until the generation of the scaffold, using as a model the *Corybacterium pseudotuberculosis* strain CIP 52.97, an intracellular patogen, which cause causeosa lymphadenitis, a disease that affects cattle and goats, with large economic loses, especially in the Northeast of Brazil, already deposited in the NCBI, with accession number CP003061. The Quality Assessment tool was used to eliminate low quality values reads (Ramos, et al. BMC Res Notes, 2011) which is based on PHRED quality values (Gordon, David, CPBioinformatics, 2004), beyond the hybrid assembly pipeline (Cerdeira *et al.*, J Microbiol Methods, 2011) that uses a set of algorithms based on De Bruijn graphs and OLC (Overlap-Layout-Consensus) and the MAUVE Assembly program to produce the scaffold. The integration of these programs is being developed in JAVA language, which will take subroutines for implementing the methods menthioned above, as well the control while the program is under implementation of the algorithms. The automation of this process will generate the high quality scaffold genome in a short time, then this genome will be submitted for the final stage of the assembly when the gaps will be closed. Until now, we did not found studies with this proposal, concerning the automation of the this process to obtain the accurate scaffold of genome.