



## **CONFIGURAÇÃO DE AMBIENTE COMPUTACIONAL PARA MONTAGEM E ANOTAÇÃO GENÔMICA: UM WORKFLOW ORIENTATIVO PARA APLICAÇÃO EM PROJETOS POR PESQUISADORES DA ÁREA DA VIDA**

Matheus Pedron Cassol (VOLUNTÁRIO), Scheila de Avila e Silva (Orientador(a))

A área da vida suscita múltiplos questionamentos que sofrem de entraves práticos e teóricos. Neste meio encontra-se o campo da bioinformática, que vêm desenvolvendo sistemas e processos que possibilitam a obtenção de novas fontes de dados e análises. Esse campo possui muitos fatores que podem afastar novos pesquisadores, como programas sem interface gráfica, uso majoritário de linhas de comando ou ainda a necessidade de hardwares com potência computacional considerável. Visando fornecer uma linha de orientação inicial dentro da bioinformática elaborou-se o presente trabalho, incentivando o uso de seus meios em mais iniciativas científicas. O método consistiu na abordagem em formato de workflow de questões bases do campo, dando enfoque à apresentação de softwares e suas aplicações em uma tabela, assim como informações para se levar em consideração durante a elaboração da pesquisa in silico. Selecionou-se alguns programas que atuam em partes distintas do processo de sequenciamento genômico buscando orientar acerca de sua instalação e uso, assim como apontar suas diferenças perante resultados. Por fim, comparou-se os resultados obtidos pelo sequenciamento de um organismo modelo, *Staphylococcus aureus*, em dois softwares, SPAdes e IDBA-UD. A avaliação da qualidade do sequenciamento foi estabelecida por testes nos programas QUASt, BUSCO e durante a anotação estrutural, pelo Augustus com apoio do BLASTP. A avaliação via QUASt retornou valores de completude em relação ao genoma referência acima de 98% para ambos testes, o que indica uma montagem confiável para o organismo em questão. Via SPAdes foi-se capaz de sequenciar com menor capacidade computacional, porém por intermédio do IDBA-UD obteve-se sequências mais contíguas. Os resultados advindos do BUSCO apresentaram um gene esperado de diferença, que se encontrou fragmentado no arquivo provindo do SPAdes e completo na versão oriunda do IDBA-UD. As proteínas e genes esperados obtidos pelo Augustus foram procuradas via BLASTP, suscitando hits, ou seja, sequências proteicas que já foram estudadas e descritas para o organismo em questão. Deste modo, pôde-se mostrar que partindo de um conhecimento básico do processo e via apropriação teórico-prática dos programas a serem utilizados pode-se obter uma montagem genômica confiável. Saber dimensionar as necessidades do projeto e traduzi-las em softwares e hardware mostra-se o passo inicial para a compreensão e capacidade de execução de processos bioinformáticos.

Palavras-chave: Bioinformática, Genoma, Sequenciamento

Apoio: UCS, outros