

**A UCS É  
PRA VOCÊ  
QUE CRIA O  
FUTURO.**



**XXIX Encontro de Jovens Pesquisadores  
e XI Mostra Acadêmica de Inovação e Tecnologia**

**De 5 a 7/10**

Local: UCS - Cidade Universitária,  
Caxias do Sul

[jovenspesquisadores.com.br](http://jovenspesquisadores.com.br)



FUNDAÇÃO  
UNIVERSIDADE DE  
CAXIAS DO SUL

**UCS**  
UNIVERSIDADE  
DE CAXIAS DO SUL

Atividade voluntária em pesquisa

**Anotação Genômica: um workflow orientativo para  
aplicação em projetos por pesquisadores da Área da  
Vida**

Autor: Matheus Pedron Cassol.  
Orientadora: Scheila de Avila e Silva.

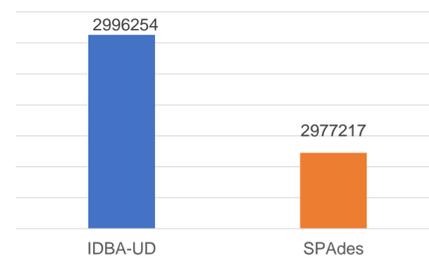


**INTRODUÇÃO / OBJETIVO**

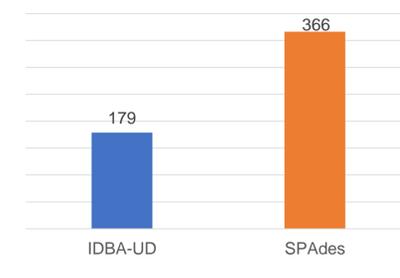
Atualmente o campo da bioinformática desenvolve ferramentas virtuais de grande valia para a execução de projetos científicos. O presente projeto visa fornecer uma linha orientativa ampla no meio bioinformático, apresentando diferentes softwares e abordando suas aplicações específicas, de modo a incentivar o uso de tais em iniciativas científicas.

**RESULTADOS E DISCUSSÃO**

Extensão total (bp)



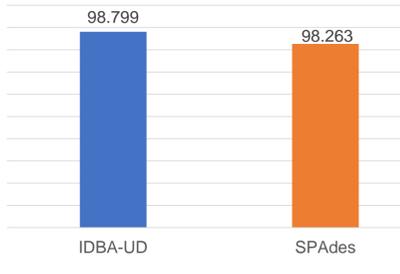
Total de contigs



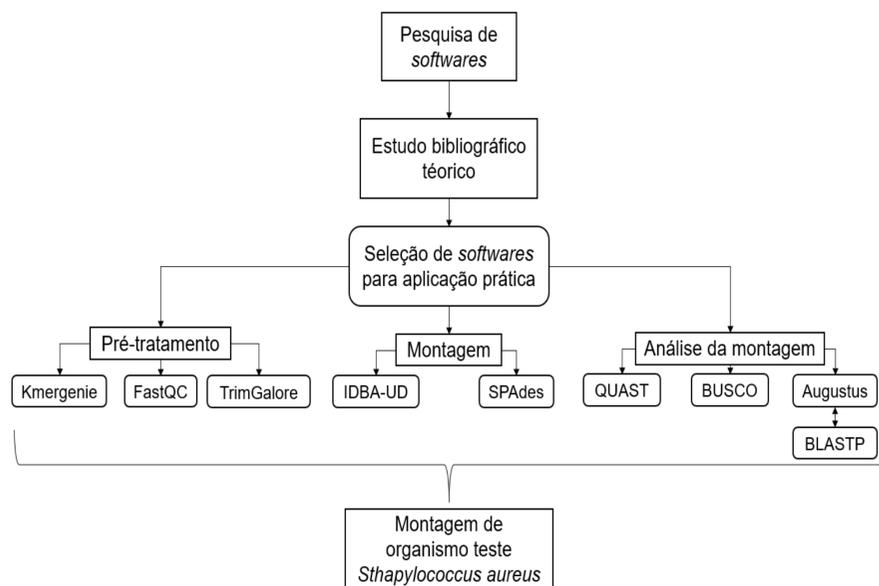
Total de BUSCOS completos



Compleitude do genoma final



**EXPERIMENTAL**



**CONCLUSÕES**

As necessidades computacionais figuram como um dos pontos de maior importância para a execução de processos bioinformáticos. Dimensionar corretamente o projeto a ser executado é um importante passo, visando buscar ferramentas e métodos apropriados para o processo.

Pôde-se ver a diferença entre softwares durante a porção prática. O montador SPAdes retornou mais sequências alinhadas em menos tempo e menor consumo de RAM, enquanto o IDBA-UD consumiu mais tempo e memória para sequenciar o mesmo arquivo inicial, apresentando maior contiguidade e completude em relação ao genoma modelo.

**RESULTADOS E DISCUSSÃO**

Via análise empregando o programa QUAST, obteve-se dados para fins de comparação entre o desempenho de ambos montadores. Assim sendo, pode-se analisar questões como completude em comparação ao genoma alvo e contiguidade da montagem. Além disso também realizou-se anotações durante o processo, a fim de abordar questões como consumo de memória RAM e intervalo de tempo necessário.

Via Augustus encontrou-se sequências proteicas e gênicas previstas para o organismo sequenciado. Após pesquisa via BLASTP encontrou-se algumas dessas sequências como já descritas para *Staphylococcus aureus*.

**REFERÊNCIAS BIBLIOGRÁFICAS**

ANDREWS, Simon et al. FastQC: a quality control tool for high throughput sequence data. 2010.

BANKEVICH, Anton et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal Of Computational Biology*. [S.L.], v. 19, n. 5, p. 455-477, maio 2012. Mary Ann Liebert Inc.

PENG, Y. et al. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, [S.L.], v. 28, n. 11, p. 1420-1428, 11 abr. 2012. Oxford University Press (OUP).

CHIKHI, R.; MEDVEDEV, P.. Informed and automated k-mer size selection for genome assembly. *Bioinformatics*, [S.L.], v. 30, n. 1, p. 31-37, 3 jun. 2013. Oxford University Press (OUP).

KELLER, Oliver; KOLLMAR, Martin; STANKE, Mario; WAACK, Stephan. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics*, [S.L.], v. 27, n. 6, p. 757-763, 6 jan. 2011. Oxford University Press (OUP).

KRUEGER, Felix. Trim galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, v. 516, p. 517, 2015.

MANNI, Mosé et al. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *arXiv preprint arXiv:2106.11799*, 2021.

MIKHEENKO, Alla; PRJIBELSKI, Andrey; SAVELIEV, Vladislav; ANTIPOV, Dmitry; GUREVICH, Alexey. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*, [S.L.], v. 34, n. 13, p. 142-150, 27 jun. 2018. Oxford University Press (OUP).

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. **Protein BLAST**: search protein databases using a protein query, search protein databases using a protein query. 2021. Disponível em: <https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>. Acesso em: 20 maio 2021.