



2023
XXXI ENCONTRO DE
**JOVENS
PESQUISADORES**

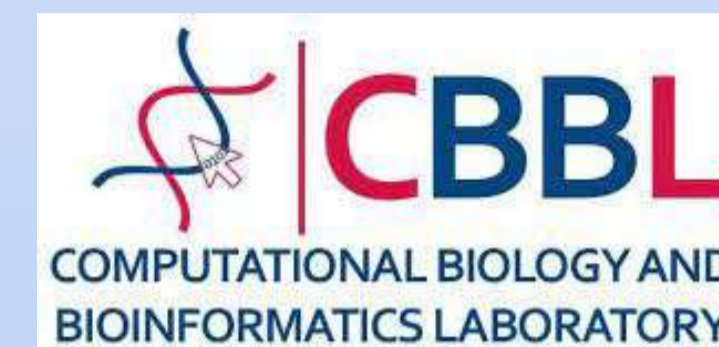


XIII Mostra Acadêmica de
Inovação e Tecnologia

Atividade voluntária em pesquisa.

Montagem genômica de levedura com potencial biotecnológico: uma abordagem *pair end* com metabolismo referência.

Autores: Matheus Pedron Cassol, Scheila de Avila e Silva.



INTRODUÇÃO / OBJETIVO

A construção de algoritmos voltados à montagem e anotação de genomas sequenciados fornece informações metabólicas de importância em diferentes áreas. Objetivou-se no presente trabalho a utilização de ferramentas bioinformáticas para efetivar a montagem e anotação de um sequenciamento *pair end* da espécie *Torulaspota delbrueckii*, com embasamento em um metabolismo de referência com relevância biotecnológica.

RESULTADOS

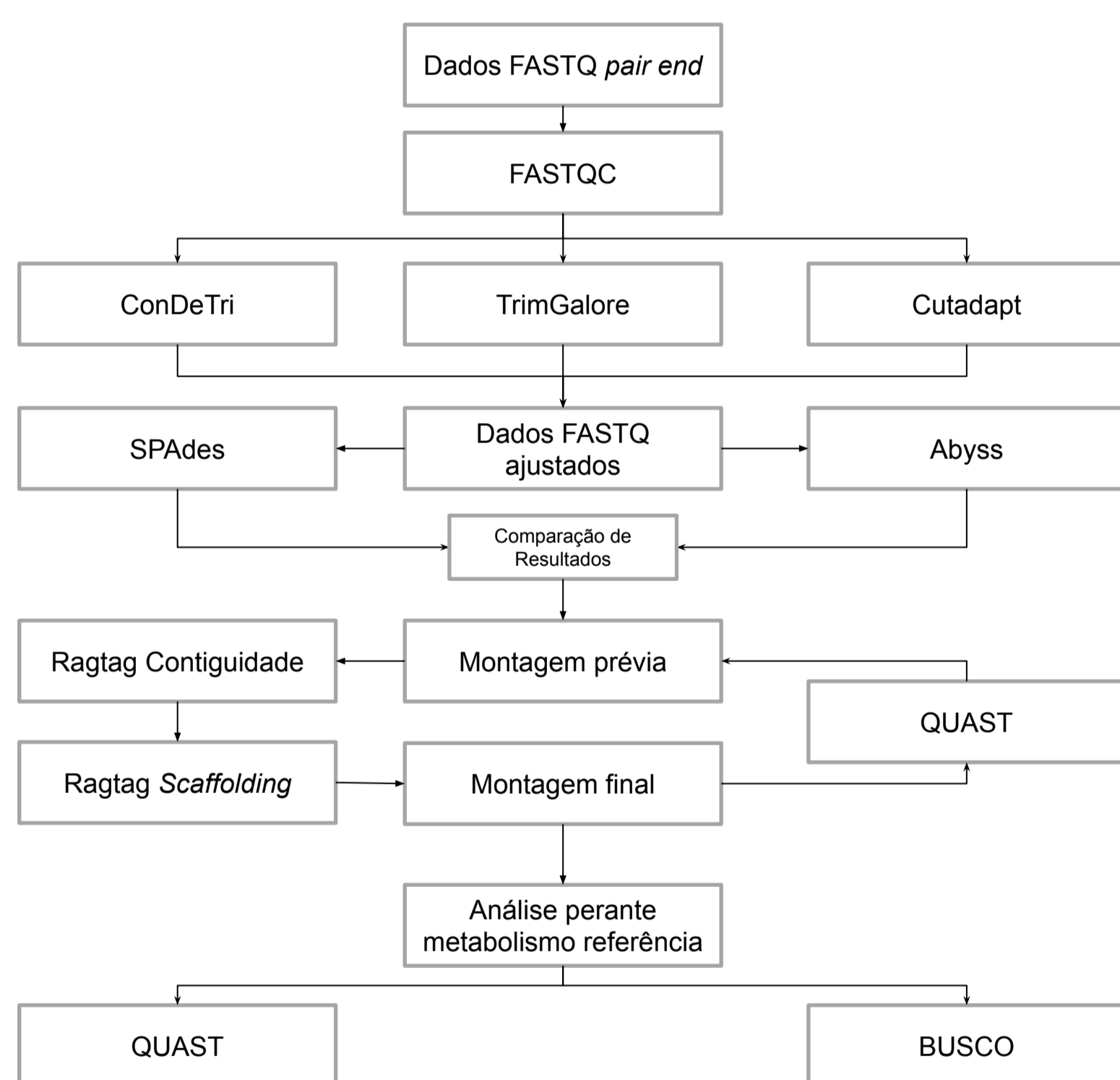
Figura 1 - Comparação de resultados entre as diferentes ferramentas de montagem utilizadas

	SPAdes	Abyss
Total de <i>contigs</i>	487	228832
Tamanho total (bp)	9302704	39371946
N50	1070550	203726
L50	4	16

Figura 2 - Dados da montagem final perante o organismo de referência

	SPAdes + RagTag
Total de <i>contigs</i>	1313
<i>Contigs</i> >= 10000 bp	23
Tamanho total (bp)	9173327
Tamanho total referência (bp)	9220678
N50	697594
Montagens equivocadas	11
Fração do genoma de referência (%)	85,600

MATERIAL E MÉTODOS



CONSIDERAÇÕES FINAIS

Pôde-se observar que ao final do processo fora obtida uma montagem confiável do genoma em estudo. Entretanto, nesta há uma presença relevante de fragmentação quando comparada ao metabolismo de referência.

Futuramente têm-se como objetivo o aprimoramento da montagem já obtida perante sua contiguidade com o auxílio de novas ferramentas bioinformáticas. Além disso, espera-se também executar o processo de anotação genômica, permitindo maior compreensão do funcionamento metabólico do indivíduo alvo.

RESULTADOS

Após o tratamento inicial dos dados pelas ferramentas de pré-processamento foram dispostas as características das duas montagens realizadas, obtidas via QUAST, em uma tabela, visando a comparação das ferramentas utilizadas, SPAdes e Abyss.

A montagem do *software* SPAdes apresentou maior fidedignidade e contiguidade, assim sendo, foi selecionada como base para o pós-processamento via RagTag. O arquivo obtido foi analisado via QUAST com base no metabolismo de referência e com a ferramenta BUSCO perante os genes típicos da classe *Saccharomycetes*. Os dados obtidos foram sintetizados em uma tabela para análise global do resultado obtido.

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.

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, Mosè 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.

MARTIN, Marcel. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnnet.Journal*, [S.L.], v. 17, n. 1, p. 10, 2 maio 2011. EMBnet Stichting.

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).

SIMPSON, Jared T.; WONG, Kim; JACKMAN, Shaun D.; SCHEIN, Jacqueline E.; JONES, Steven J.M.; BIROL, Inanc. ABYSS: a parallel assembler for short read sequence data. *Genome Research*, [S.L.], v. 19, n. 6, p. 1117-1123, 27 fev. 2009. Cold Spring Harbor Laboratory.

SMEDS, Linnéa; KÜNSTNER, Axel. ConDeTri - A Content Dependent Read Trimmer for Illumina Data. *Plos One*, [S.L.], v. 6, n. 10, p. e26314, 19 out. 2011. Public Library of Science (PLoS).