

# Genome assembly and variant analysis of two *Saccharomyces cerevisiae* strains isolated from stingless bee pollen

Eduardo Luís Menezes de Almeida<sup>1\*</sup>, Wendel Batista da Silveira<sup>1</sup>, Luciano Gomes Fietto<sup>2</sup>, Weyder Cristiano Santana<sup>3</sup>, Monique Renon Eller<sup>4</sup>

<sup>1</sup>Laboratory of Microbial Physiology, Department of Microbiology, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil.

<sup>2</sup>Department of Biochemistry, Universidade Federal de Viçosa, Viçosa, Brazil.

<sup>3</sup>Department of Entomology, Universidade Federal de Viçosa, Viçosa, Brazil.

<sup>4</sup>Department of Food Technology, Universidade Federal de Viçosa, Viçosa, Brazil.

\*Corresponding author. E-mail: eduardo.menezes@ufv.br

Products from stingless bees, such as pollen, are rich reservoirs of microbial diversity, including yeasts with fermentative potential. Previously, we isolated two *Saccharomyces cerevisiae* strains, JP14 and IP9, from Jataí (*Tetragonisca angustula*) and Iraí (*Nannotrigona testaceicornis*) bees, respectively, aiming at mead production. Both strains presented great osmotic and sulfite tolerance and ethanol production, although they have a high free amino nitrogen demand. Herein, we sequenced, assembled, and annotated their genomes and evaluated the variants compared to *S. cerevisiae* S288c (reference). After filtering, we assembled the reads using SPAdes and corrected the contigs and assembled the scaffolds using RagTag. QUAST and BUSCO assessed the quality of assemblies. We transferred the annotation of coding sequences from S288c using LiftOff and annotated the mitochondrial genome using MITOS2. We used nQuire to predict the ploidy of IP9 and JP14, and BWA-MEM to align their reads with S288c. Freebays called the genomic variants, snpEff annotated them, and Control-FREEC detected copy number variants (CNVs). Finally, we conducted a Gene Ontology (GO) functional enrichment analysis of mutations and CNVs. The final assembly of IP9 and JP14 presented BUSCO scores above 99.5%, N50 above 888,000 bp, and more than 6430 protein-coding genes. Although the mitochondrial genomes of IP9 and JP14 were rearranged, they presented the same gene content as S288c. Meanwhile, the genes with mutations of high impact (e.g., indels, gain of stop codon) for both yeasts were enriched for transmembrane transport, electron transfer, oxidoreductase, heme binding, fructose, mannose, and glucose transport, activities related to the respiratory chain and sugar metabolism. The regions with CN gains in IP9 were related to genes related to sugar transport and cell morphogenesis; in JP14, genes were enriched for disaccharide metabolism and transport, response to reactive oxygen species, and polyamine transport. On the other hand, IP9 presented CN losses related to disaccharide, thiamine, and aldehyde metabolism, while JP14 presented depletions related to disaccharide, oligosaccharide, asparagine, and aspartate metabolism. In summary, we successfully assembled and annotated the genome of JP14 and IP9 and found a high selective pressure for sugar and nitrogen metabolism and stress response, consistent with their isolation source and fermentative properties.

**Keywords:** Comparative genomics; Beverage production; Bioinformatics.

## Montagem do genoma e análise de variantes de duas linhagens de *Saccharomyces cerevisiae* isoladas do pólen de abelhas sem ferrão

*Saccharomyces cerevisiae* JP14 e IP9 foram isoladas de pólen, apresentando tolerância osmótica e ao sulfito, alta produção de etanol e demanda nutricional. Seus genomas foram sequenciados, montados e anotados com alta qualidade. As linhagens apresentaram número de cópias (NC) enriquecidos para genes relacionados a transporte transmembrana, transferência de elétrons e metabolismo de açúcares. IP9 apresentou NC enriquecidas para transporte de açúcares e morfogênese celular, e JP14 para metabolismo de dissacarídeos, respostas ao estresse e transporte de poliaminas. Isso reflete as alterações metabólicas consistentes com as adaptações necessárias à sua sobrevivência e metabolismo na fonte de isolamento.

**Palavras-chave:** Genômica comparativa; Produção de bebidas; Bioinformática.

Acknowledgements: Ministério da Ciência, Tecnologia e Inovações (MCTI), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq Finance Code 140538/2021-6), Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES) – Finance Code 001.