

Recovery of 447 Eukaryotic bins reveals major challenges for Eukaryote genome reconstruction from metagenomes

Joao Saraiva¹, Alexander Bartholomäus², Rodolfo Toscan¹, Petr Baldrian³, and Ulisses da Rocha¹

¹Helmholtz-Centre for Environmental Research - UFZ

²Helmholtz Centre Potsdam German Research Centre for Geosciences

³Institute of Microbiology of the ASCR

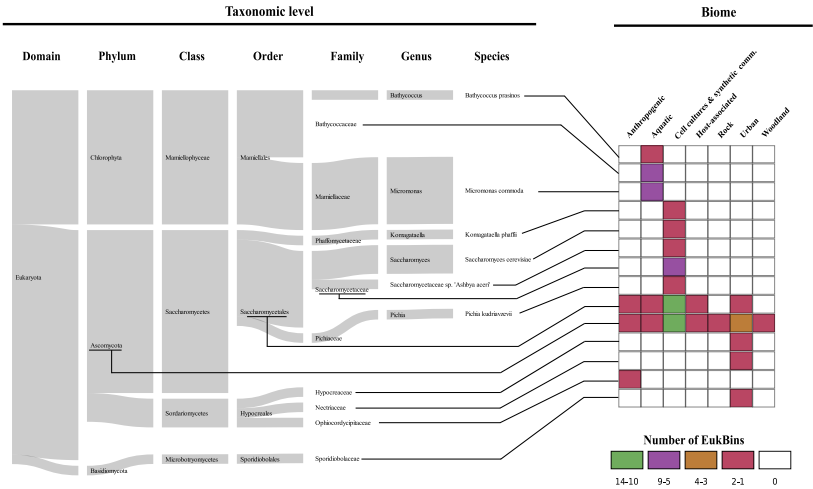
April 8, 2022

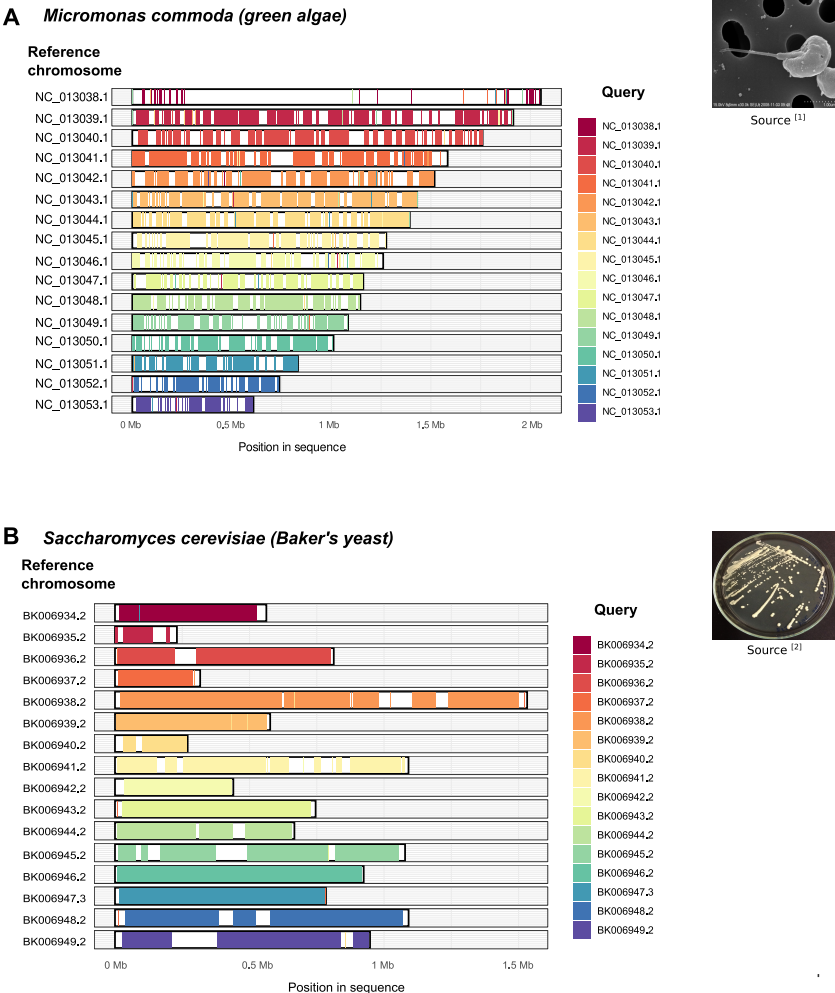
Abstract

An estimated 8.7 million eukaryotic species exist on our planet. However, recent tools for taxonomic classification of eukaryotes only dispose of 734 reference genomes. As most Eukaryotic genomes are yet to be sequenced, the mechanisms underlying their contribution to different ecosystem processes remain untapped. Although approaches to recover Prokaryotic genomes have become common in genome biology, few studies have tackled the recovery of Eukaryotic genomes from metagenomes. This study assessed the reconstruction of Eukaryotic genomes using 215 metagenomes from diverse environments using the EukRep pipeline. We obtained 447 eukaryotic bins from 15 classes (e.g., Saccharomycetes, Sordariomycetes, and Mamiellophyceae) and 16 orders (e.g., Mamiellales, Saccharomycetales, and Hypocreales). More than 73% of the obtained eukaryotic bins were recovered from samples whose biomes were classified as host-associated, aquatic and anthropogenic terrestrial. However, only 93 bins showed taxonomic classification to (9 unique) genera and 17 bins to (6 unique) species. A total of 193 bins contained completeness and contamination measures. Average completeness and contamination were 44.64% ($\sigma=27.41\%$) and 3.97% ($\sigma=6.53\%$), respectively. *Micromonas commoda* was the most frequent taxa found while *Saccharomyces cerevisiae* presented the highest completeness, possibly resulting from a more significant number of reference genomes. However, mapping eukaryotic bins to the chromosomes of the reference genomes suggests that completeness measures should consider both single-copy genes and chromosome coverage. Recovering eukaryotic genomes will benefit significantly from long-read sequencing, intron removal after assembly, and improved reference genomes databases.

Hosted file

02_Saraiva_et_al_Molec_Ecology_Resources_v7_updated.pdf available at <https://authorea.com/users/475051/articles/564601-recovery-of-447-eukaryotic-bins-reveals-major-challenges-for-eukaryote-genome-reconstruction-from-metagenomes>





Hosted file

Saraiva_et_al_Table_1.docx available at <https://authorea.com/users/475051/articles/564601-recovery-of-447-eukaryotic-bins-reveals-major-challenges-for-eukaryote-genome-reconstruction-from-metagenomes>