# ATCC Genome Portal: Fungal Genome Assembly Improvements and Mitochondrial DNA Recovery

ATCC®
Credible leads to Incredible®

Emily A. White, MS; Joseph R. Petrone, PhD; Scott V. Nguyen; PhD, Nikhita P. Puthuveetil, MS; David A. Yarmosh, MS; Amy L. Reese, MS; John Bagnoli, BS; Briana Benton, BS; Jonathan L. Jacobs, PhD
ATCC, Manassas, VA 20110

## Abstract

The utility of genomic data is constrained by the tools, assumptions, and reference quality available at the time the data is generated; consequently, many datasets contain biological signals that remain obscured until advances in analytical methods enable strategic reanalysis. Fungal mitochondrial genomes are a case in point: high-quality references now include 3,131 JGI-annotated and 892 complete RefSeq mitogenomes, yet 6,467 undeclared mitogenomes remain embedded in GenBank nuclear assemblies.[1] Mitogenomes further complicate analysis through extreme size variation (~2 kb to >200 kb), mixed architectures, absence of universal marker genes, and nuclear insertions (NUMTs).[1] Here, we compare 37 authenticated ATCC® Genome Portal (AGP) fungal assemblies against strain-matched RefSeq genomes using standard quality metrics. Because mitochondrial representation remains inconsistent across public databases, we assessed mitogenome integration in a subset of ATCC® assemblies using curated RefSeq references and explore the impacts of implementing a targeted workflow to separate mitochondrial and nuclear reads prior to assembly.

## Methods

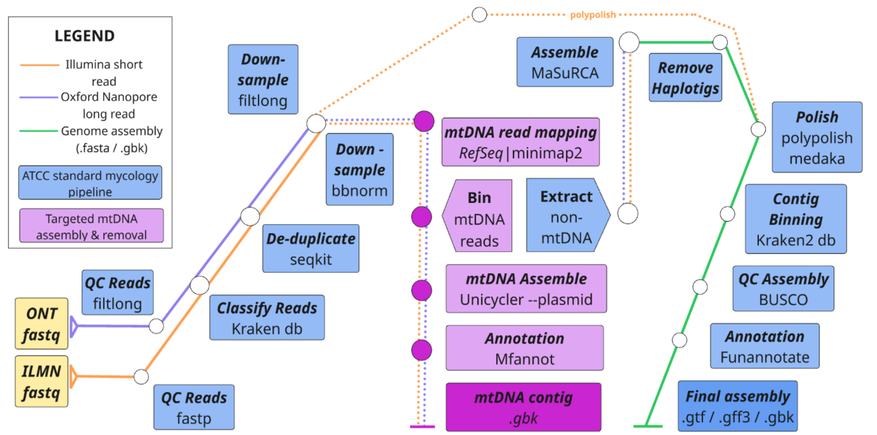


**Figure 1: AGP fungal genome assembly pipeline highlighting targeted mitochondrial extraction (non-standard workflow).** A de novo hybrid assembly workflow (blue) with a derivative mitochondrial extraction pipeline (pink). After downsampling, Illumina and ONT reads (dotted lines) branch to support parallel processing: mitochondrial reads are identified via alignment to strain-matched RefSeq mitogenomes, extracted, and assembled independently, while mitochondrial-mapped reads are removed from nuclear read sets. Following mtDNA extraction, Illumina and ONT reads are combined in a hybrid MaSuRCA assembly. A secondary dotted orange branch from the initial pre-processed Illumina reads represents downstream polishing (e.g., Polypolish).

## Results: Mitogenome

**Table 1: *Rhodotorula mucilaginosa* (ATCC® 2503™) assembly contiguity and BUSCO completeness metrics before and after mtDNA removal.**

| Assembly Version | Contigs (n) | N50 (nt) | Genome Size | Completeness |
|---|---|---|---|---|
| ATCC® 2503™ [standard] | 24 | 1,614,595 | 20.4Mb | 92.70% |
| ATCC® 2503™ [targeted] | 28 | 1,270,344 | 20.6 Mb | 96.30% |

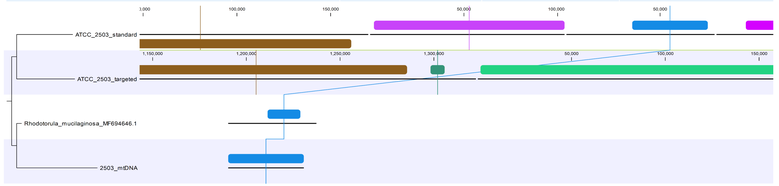


**Figure 2: CLC whole-genome alignment of *Rhodotorula mucilaginosa* (ATCC_2503) assemblies to the RefSeq mitogenome (MF69464.1; 47,023 bp).** Homology is color-coded. The targeted assembly excludes the mitochondrial contig (blue), and the de novo mtDNA assembly (bottom track; 40,289 bp) shares 99.74% ANI with the reference.

## Results: Strain Comparisons

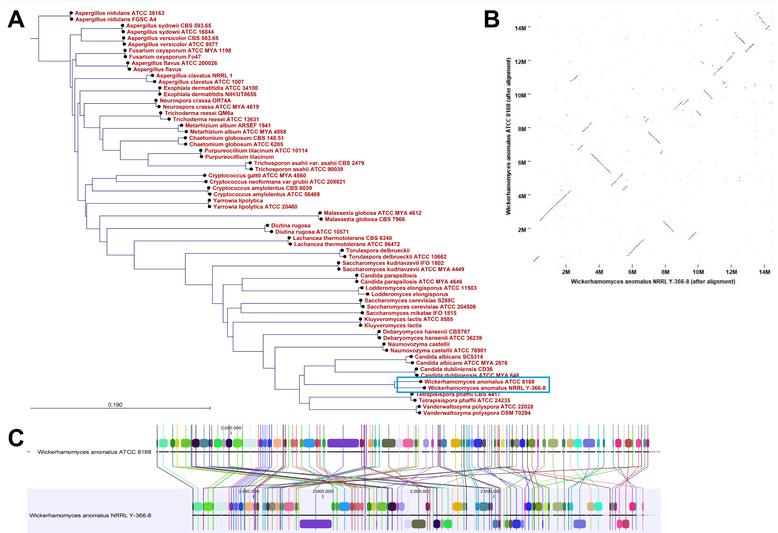


**Figure 3: Comparison of strain-matched AGP and RefSeq fungal assemblies.** (A) K-mer tree showing close clustering for most pairs; outliers suggest contamination, misassembly, or quality differences. (B) CLC Dot plot and (C) whole-genome alignment for *W. anomalus* (blue box in [A]) show structural differences despite >97% ANI and ~91% gene-level identity.

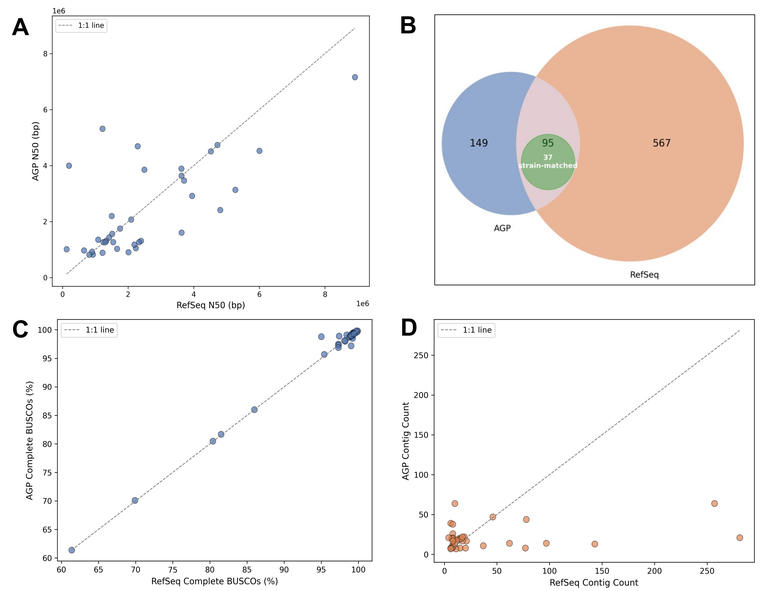


**Figure 4: Comparison of AGP and strain-matched RefSeq fungal assembly (n = 37) quality metrics.** (A) N50 values. (B) Species representation Venn diagram. (C) BUSCO completeness (%; eukaryota_odb12). (D) Contig count.

## Conclusions

- ATCC® assemblies exhibit reduced fragmentation relative to strain-matched RefSeq genomes while maintaining comparable BUSCO completeness.
- Read-recruitment against taxa-specific mitochondrial references, followed by binned reassembly, further improves nuclear genome quality by removing embedded mitochondrial sequences.
- These findings support continued refinement of fungal assembly workflows via provenance-aware, reference-guided approaches.

Learn more about the ATCC® Genome Portal

## References

1. Ahrendt SR, et al. Comparative mitogenomics of kingdom Fungi - evolutionary insights and metagenomic applications. Nucleic Acids Res 54(2): gkaf1419, 2026. PubMed: 41533582