Poster Number: AEM-SUN-641
Date: June 7, 2026

Generation and Annotation of *Yarrowia lipolytica* Genome Assemblies

ATCC®
Credible leads to Incredible®
snguyen@atcc.org

Scott V. Nguyen, PhD; Joseph Petrone, PhD; Nikhita Puthuveetil, MS; David Yarmosh, MS; Amy Reese, MS; Emily White, MS; Corina Tabron, MS; Jade Kirkland, BS; Kaitlyn Bentley, MS; James Duncan, MS; Robert Marlow, BS; Stephen King, MS; Noah Wax, MS; Ana Fernandes, BS; John Bagnoli, BS; Briana Benton, BS; Jonathan L. Jacobs, PhD
ATCC, Manassas, VA 20110

Introduction

Yarrowia lipolytica is an oleaginous yeast with exceptional metabolic flexibility and a strong capacity for fatty acid accumulation. While *Saccharomyces cerevisiae* remains the most used chassis in biotechnology, *Y. lipolytica* offers distinct advantages under conditions that inhibit many conventional yeasts. Numerous *Y. lipolytica* strains are preserved in public biorepositories, including strains of applied and academic relevance. Several of these have been sequenced, and reference-quality genome assemblies are available through public genomic databases. In this study, we selected *Y. lipolytica* strains with broad applied, academic, and taxonomic significance from a publicly accessible *Yarrowia* collection. Genomes were sequenced using both Illumina and Oxford Nanopore technologies and annotated with a custom database, yielding high-quality, fully annotated assemblies. These assemblies were deposited in a publicly accessible genomic resource and cross-linked to the originating biological materials to maintain strain-genome traceability. This dataset includes the first phased diploid assembly of the *Y. lipolytica* type strain (ATCC® 18942™) and an improved assembly relative to the current RefSeq reference. By providing accurate and traceable genomic resources, this work supports ongoing research into *Y. lipolytica* biology and its development as a platform for microbial production and systems biology research.

Methods

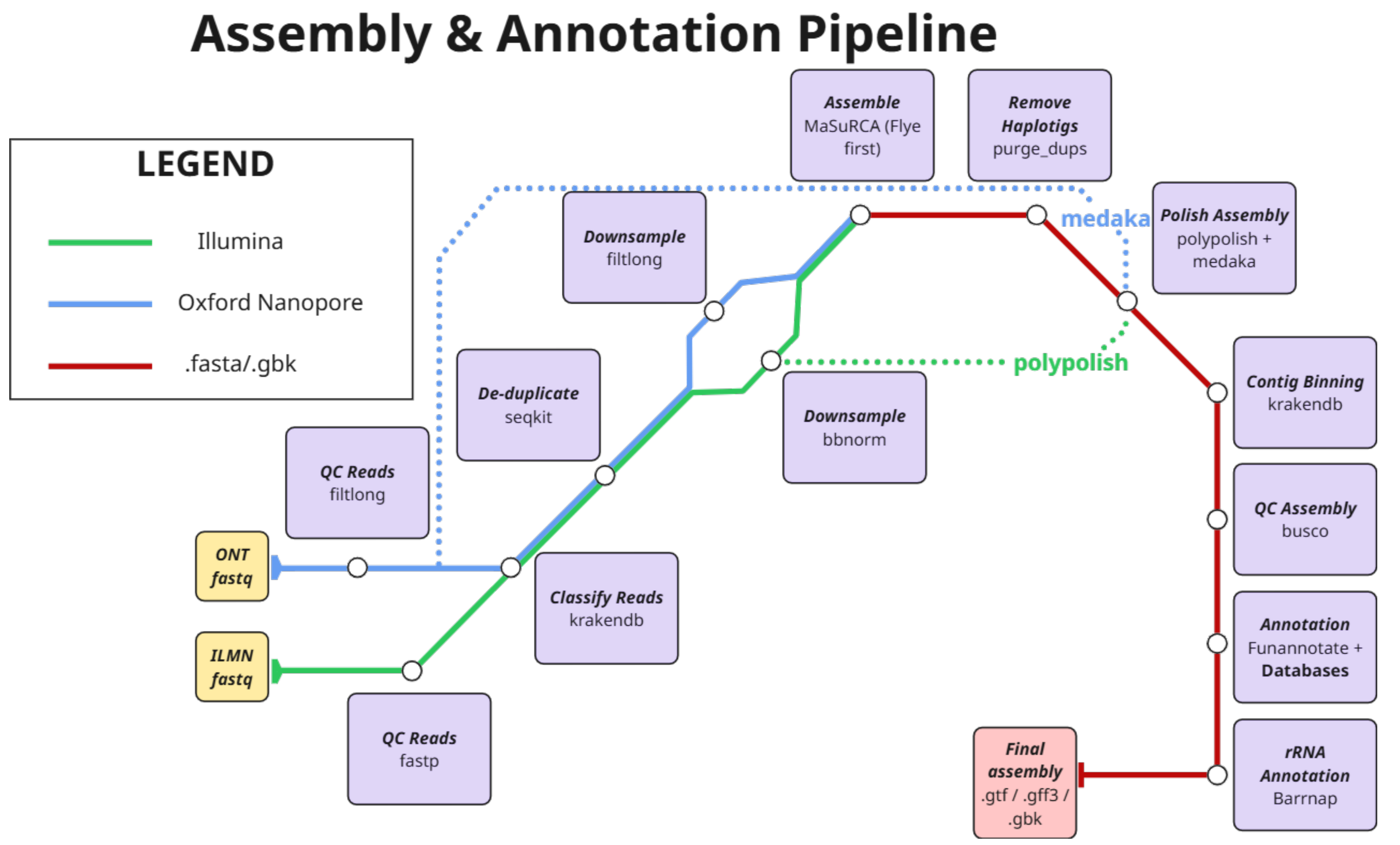


Figure 1: Genome assembly and annotation pipeline. A hybrid short- and long-read approach is used to assemble high-quality genomes of *Y. lipolytica* at ATCC®. This strategy results in higher quality and more contiguous genomes compared to those found in NCBI.

Results

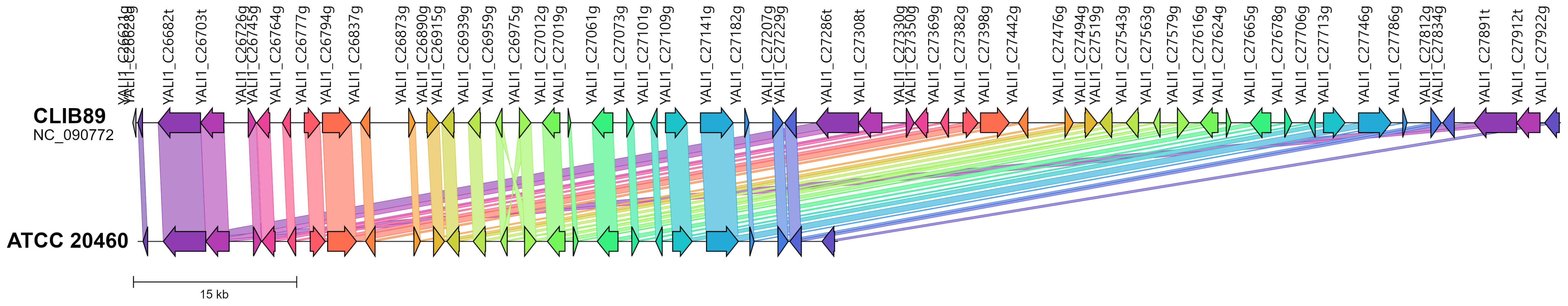


Figure 2: CLIB89 reference compared to ATCC® 20460™. The reference strain in NCBI for *Y. lipolytica* is CLIB89, also known as W29, and is available at ATCC® as ATCC® 20460™. A 54-kb repeat on chromosome C in CLIB89 that was described in the CLIB89 manuscript was not in the ATCC® 20460™ assembly.

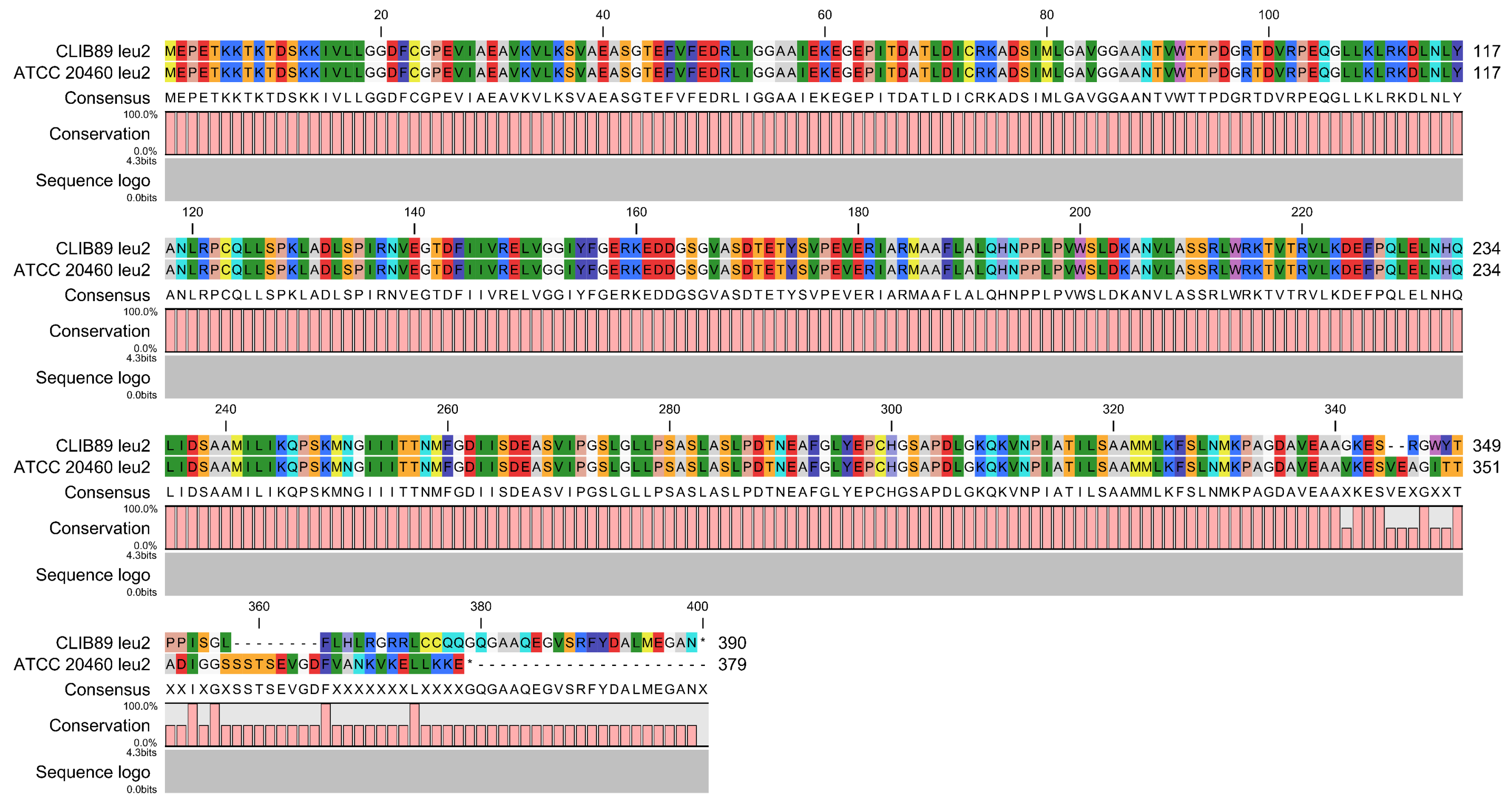


Figure 3: Apparent C-terminal discrepancy in the LEU2 reference sequence for CLIB89. Protein sequence alignment of LEU2 from the CLIB89 reference genome and ATCC® 20460™. The *leu2* gene encodes β-isopropylmalate dehydrogenase, a key genetic marker. The C-terminus discrepancy is consistent with a potential misassembly affecting the 3′ end of the *leu2* gene locus in the CLIB89 reference. Read mapping for ATCC® 20460™ supports the ATCC® 20460™ *leu2* gene sequence and the inferred C-terminal structure of LEU2.

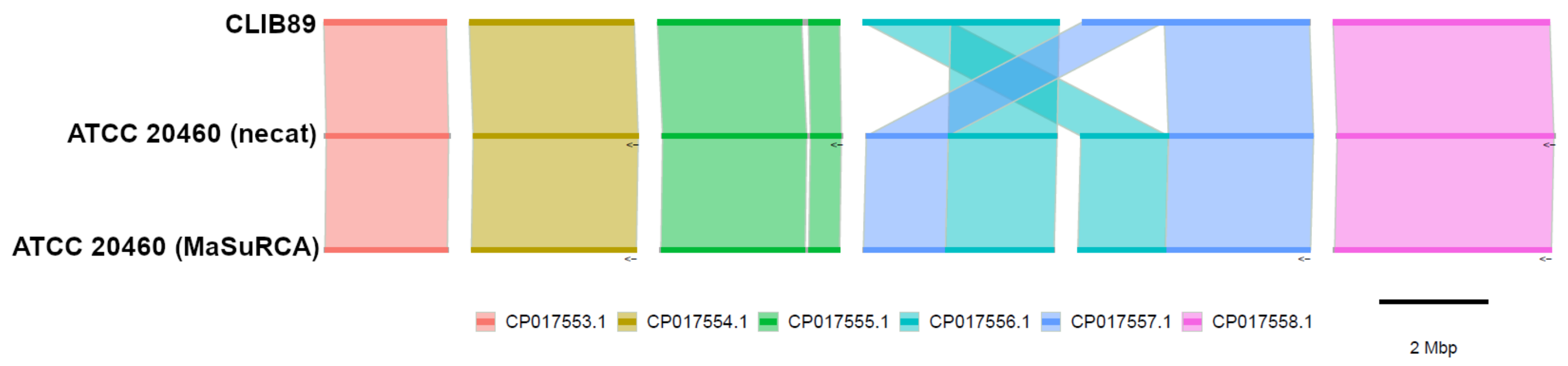
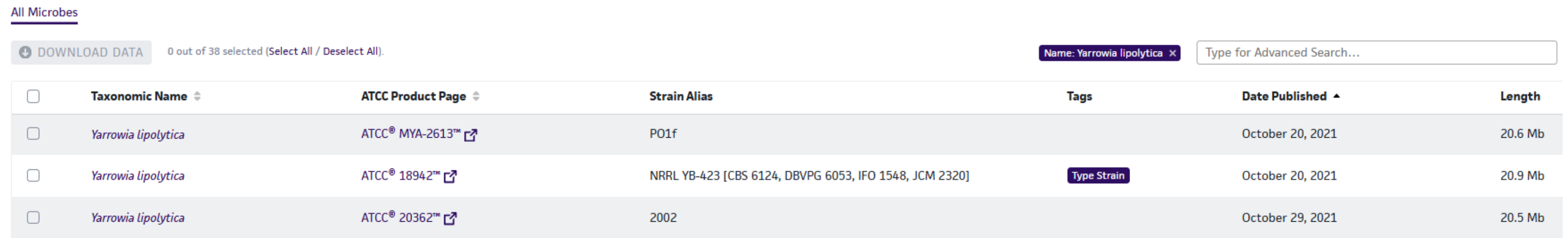


Figure 4: Whole-genome synteny between the CLIB89 reference and *de novo* assemblies of ATCC® 20460™. Synteny plots generated using ntSynt comparing the CLIB89 reference (top) with independent *de novo* assemblies of ATCC® 20460™ using NECAT (middle) and MaSuRCA (bottom). Colored blocks correspond to homologous chromosomal segments with connecting ribbons indicating conserved synteny. Both ATCC® 20460™ assemblies show high overall synteny with rearrangements visible in specific regions to CLIB89.



0 out of 38 selected (Select All / Deselect All).

☐	Taxonomic Name	ATCC Product Page	Strain Alias	Tags	Date Published	Length
☐	*Yarrowia lipolytica*	ATCC® MYA-2613™	PO1f		October 20, 2021	20.6 Mb
☐	*Yarrowia lipolytica*	ATCC® 18942™	NRRL YB-423 [CBS 6124, DBVPG 6053, IFO 1548, JCM 2320]	Type Strain	October 20, 2021	20.9 Mb
☐	*Yarrowia lipolytica*	ATCC® 20362™	2002		October 29, 2021	20.5 Mb

Figure 5: ATCC® Genome Portal. 38 hybrid genome assemblies of *Y. lipoytica* are in the ATCC® Genome Portal (https://genomes.atcc.org/), including the diploid assembly of the type strain and PO1f, an important industrial strain.

Conclusions

- High-quality hybrid short- and long-read sequencing resolves structural discrepancies present in widely used *Y. lipolytica* reference genomes, improving accuracy at both gene and chromosome scales.
- Linking curated genome assemblies to authenticated biological materials in a public biorepository ensures long-term data provenance and reproducibility.
- The ATCC® Genome Portal enables community reuse of traceable, high quality reference genomes to support biotechnology and systems biology research.

Learn more about the ATCC® Genome Portal

References

1. Crill JE II, et al. Complete diploid genome of the type strain *Yarrowia lipolytica* YB-423 (ATCC 18942™). J Ind Microbiol Biotechnol 53: kuag002, 2026. PubMed: 41504366
2. Nguyen SV, et al. The ATCC genome portal: 3,938 authenticated microbial reference genomes. Microbiol Resour Announc 13(2): e0104523, 2024. PubMed: 38289057
3. Benton B, et al. The ATCC Genome Portal: Microbial Genome Reference Standards with Data Provenance. Microbiol Resour Announc 10(47): e0081821, 2021. PubMed: 34817215
4. Yarmosh DA, et al. Comparative Analysis and Data Provenance for 1,113 Bacterial Genome Assemblies. mSphere 7(3): e0007722, 2022. PubMed: 35491842