POM-SUNDAY-1207
June 16, 2024

The ATCC® Genome Portal: Authenticated Microbial Reference Genomes

Joseph R. Petrone, PhD; Nikhita P. Puthuveetil, MS; David A. Yarmosh, MS; Amy L. Reese, MS; Corina Tabron, MS; Jade Kirkland, BS; Kaitlyn Gaffney, MS; Noah Wax, MS; James Duncan, MS; Robert Marlow, BS; Stephen King, MS; Scott Nguyen, PhD; John Bagnoli, BS; Briana Benton, BS; Jonathan L. Jacobs, PhD | ATCC, Manassas, VA 20110

Background

The ATCC® Genome Portal was developed as part of an initiative to produce high-quality reference genomes for the ATCC® microbial collection. Currently, the portal includes fully authenticated and annotated genome assemblies for over 4,500 microbes, encompassing 3,828 bacterial, 362 viral, 306 fungal, and 4 protist genomes. The ATCC® Genome Portal (AGP), which warehouses these data, is meticulously curated and links all the metadata and sequencing data to physical materials in ATCC®'s collection. As the sole ISO 9001–compliant reference database for ATCC® microbial strains, the portal ensures full end-to-end data provenance and authentication. Accessible for research purposes, the ATCC® Genome Portal and its data can be explored online (https://genomes.atcc.org) or through an authenticated REST-API.

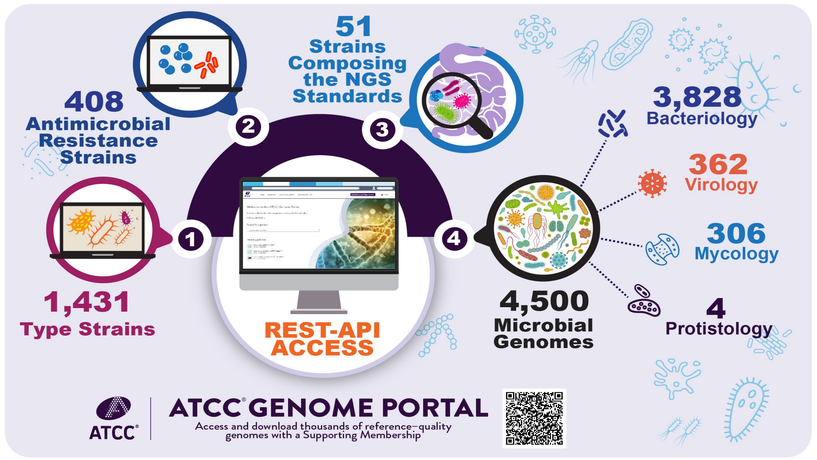

Figure 1: Reference genomes available through the ATCC® Genome Portal. The statistics are current as of May 1, 2024.

AGP Supporting Membership

Table 1: ATCC® Genome Portal Supporting Membership plans

	Free Plan	Individual Member	Research Member	Institutional Member
View organism and genome metadata, assemblies, and annotations	✓	✓	✓	✓
Search for genes of interest	✓	✓	✓	✓
Download genome assemblies and annotations	Only for purchased products	✓	✓	✓
Access the REST API	Not available	✓	✓	✓
Compare your sequencing data to ATCC® genomes with Discrepancy Reports	Fee for each report	12 free reports per year	60 free reports per year	Inquire for details
Members with full access	0	1	5	Unlimited

Methodology

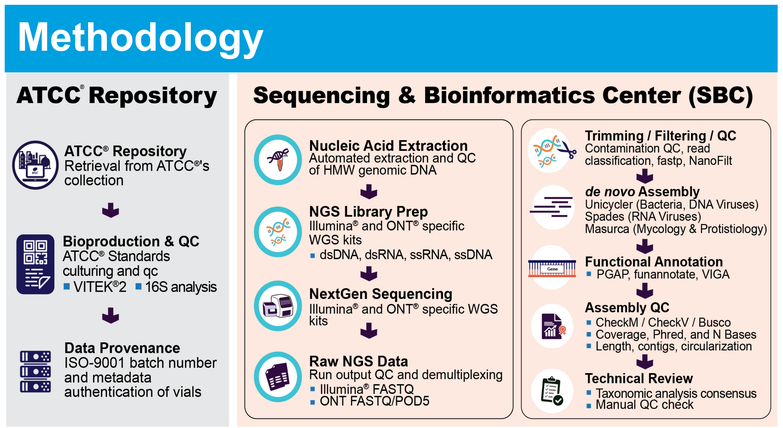

Figure 2: Standard pipeline for end-to-end genome portal publication.

Authenticated ATCC® Genomes

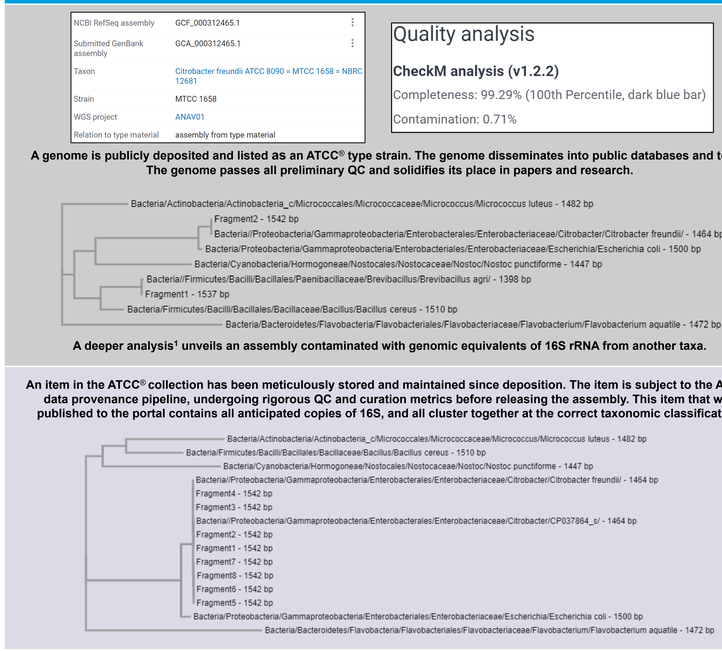

Figure 3: The genome for an ATCC® item deposited into public databases that is not authenticated and contains critical errors as compared to an ATCC®-curated assembly for that same item.

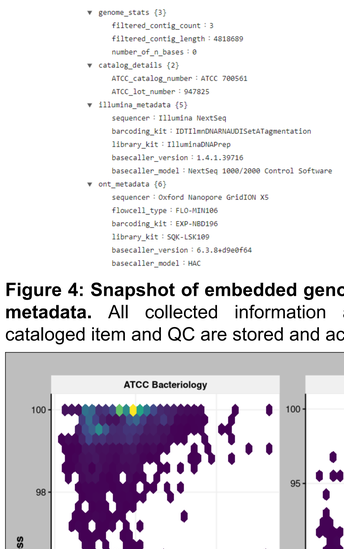

Figure 4: Snapshot of embedded genome JSON metadata. All collected information about the cataloged item and QC are stored and accessible.

Table 2: Additional JSON fields available to access.

Additional JSON Fields	Values
Input Reads Summary	6
Annotations Summary	8
AMR Summary	4
Genome Provider	6
NGS Read QC (ILM/ONT)	4/4
Assembly QC	4
Product Page Info	>18

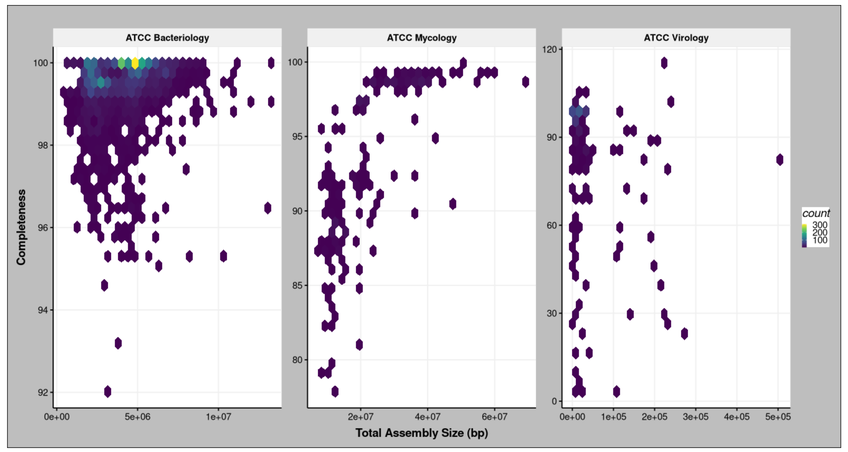

Figure 5: Completeness metrics of the ATCC® Genome Portal. CheckM, CheckV, and Busco were used for Bacteriology, Virology, and Mycology products, respectively.[2,3]

Conclusions

- Public genome databases lack the authentication and traceability that is provided by the ATCC® Genome Portal. This gap in data provenance potentially complicates downstream bioinformatics applications and research objectives.[4]
- ATCC® is producing ultra-high quality reference genomes for all microbial species in our collection and providing research-use only access to these data via the ATCC® Genome Portal.
- All genomes from the collection follow a standardized protocol and rigorous QC from deposition and extraction to sequencing and assembly and are released quarterly to the portal.[5]

References

1. Lee I, et al. ContEst16S: an algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences. In J Syst Evol Microbiol 67(6): 2053-2057, 2017.
2. Parks DH, et al. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25(7): 1043-1055, 2015.
3. Simão FA, et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19): 3210-3212, 2015.
4. Yarmosh DA, et al. Comparative analysis and data provenance for 1,113 bacterial genome assemblies. mSphere 7(3): e00077-22, 2022.
5. Nguyen SV, et al. The ATCC Genome Portal: 3,938 authenticated microbial reference genomes. Microbiol Resour Announc 13(2): e0104523, 2024.

Discover The ATCC® Genome Portal REST-API

ATCC 10801 University Boulevard, Manassas, Virginia 20110-2209 **Phone:** 800.638.6597 **Email:** sales@atcc.org **Web:** www.atcc.org