MP-04
July 29, 2025

The ATCC Genome Portal: The Genomics Database of a 100-year-old Culture Collection

Scott V Nguyen, PhD; Ana Fernandes, BS; John Bagnoli, BS; Briana Benton, BS; Jonathan L Jacobs, PhD | ATCC, Manassas, VA 20110

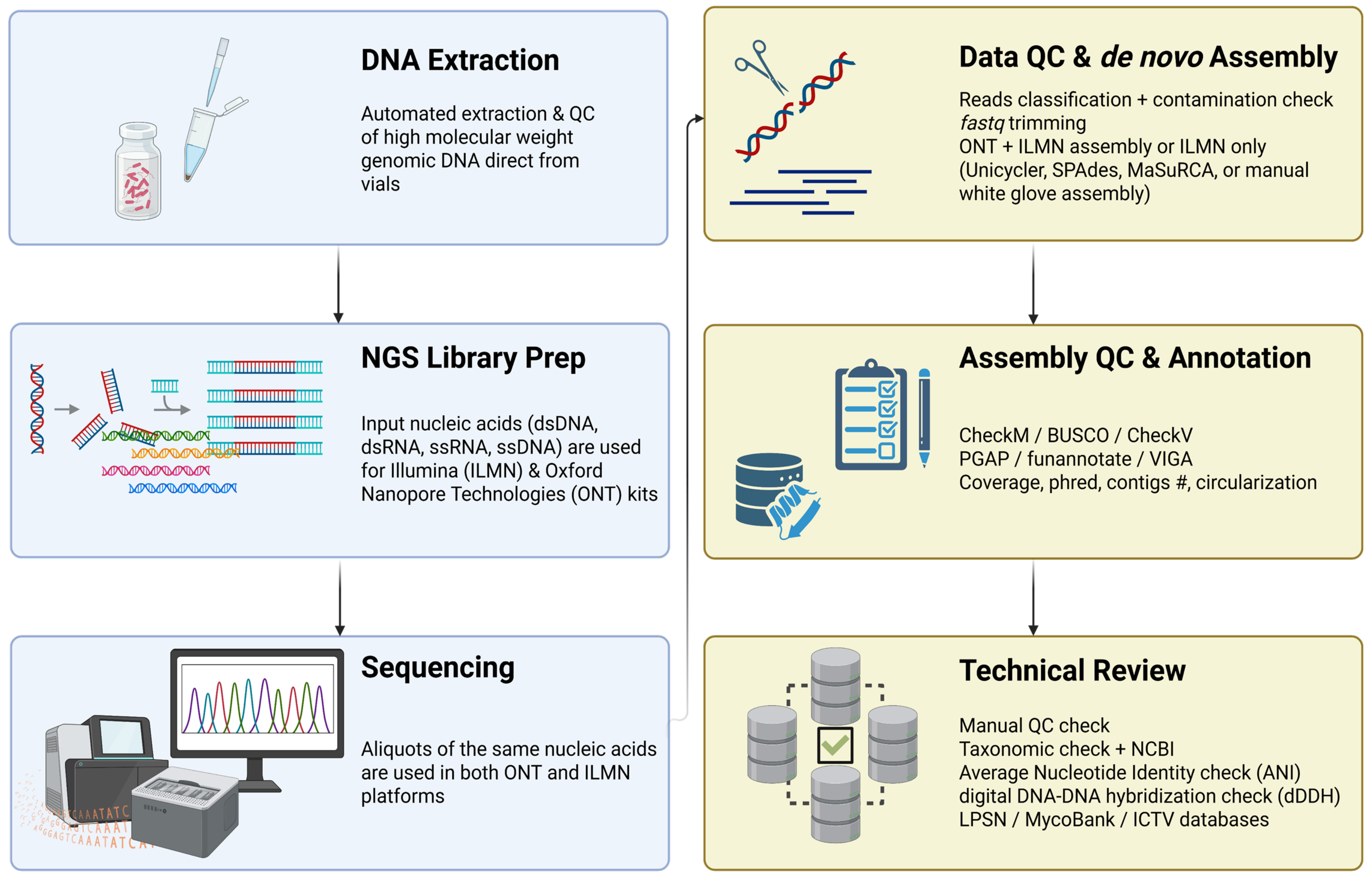
Introduction

The ATCC® Genome Portal (AGP, https://genomes.atcc.org) is a reference database of authenticated and traceable whole-genome sequence assemblies generated directly from strain vials from one of the largest and most diverse culture collections in the world. As a non-profit biorepository, thousands of scientists and institutions have deposited numerous microbes, fungi, protists, viruses, cell lines, plasmids, and patented strains into ATCC over the past century.

Purpose:

The AGP provides a resource to the scientific community as many genomes that are in public genomic databases may represent sequences from strains that have been passed between laboratories, are mislabeled, or have accumulated genetic drift arising from laboratory domestication. Thousands of foodborne pathogens and food related microbes have been deposited into the collection, with isolates representing over a century of research. Currently, there are over 5,750 reference quality genomes of prokaryotes, fungi, viruses, and protists available in the AGP with new additional assemblies uploaded each quarter.

Methods



DNA Extraction

Automated extraction & QC of high molecular weight genomic DNA direct from vials

NGS Library Prep

Input nucleic acids (dsDNA, dsRNA, ssRNA, ssDNA) are used for Illumina (ILMN) & Oxford Nanopore Technologies (ONT) kits

Sequencing

Aliquots of the same nucleic acids are used in both ONT and ILMN platforms

Data QC & *de novo* Assembly

Reads classification + contamination check
fastq trimming
ONT + ILMN assembly or ILMN only (Unicycler, SPAdes, MaSuRCA, or manual white glove assembly)

Assembly QC & Annotation

CheckM / BUSCO / CheckV
PGAP / funannotate / VIGA
Coverage, phred, contigs #, circularization

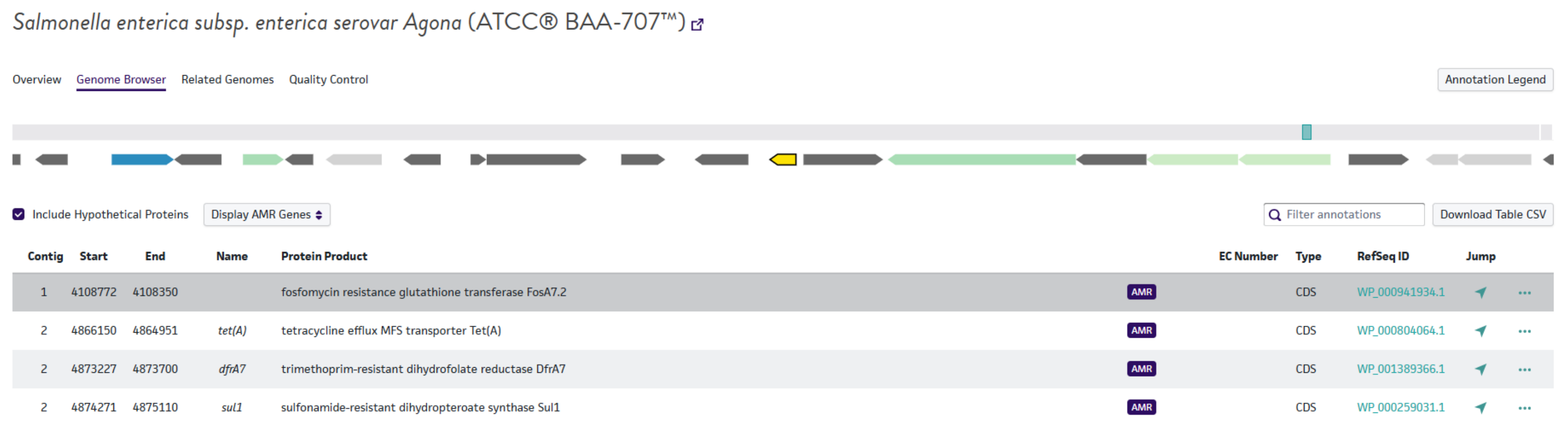
Technical Review

Manual QC check
Taxonomic check + NCBI
Average Nucleotide Identity check (ANI)
digital DNA-DNA hybridization check (dDDH)
LPSN / MycoBank / ICTV databases

Figure 1: Standard pipeline for data provenance and sequencing of strains at ATCC. Created with BioRender.com.

Results

An Authenticated Reference Database for a Non-profit Culture Collection

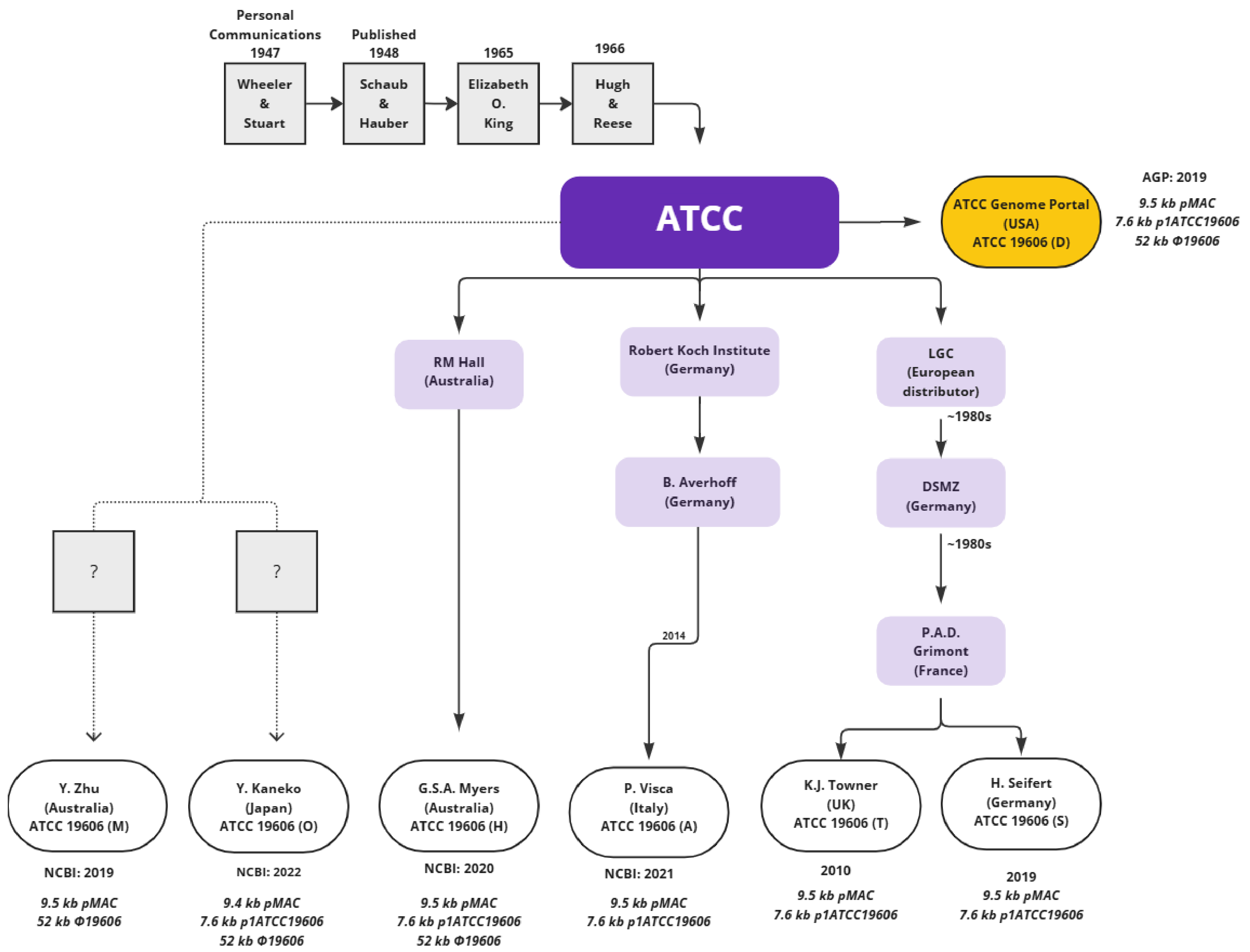


Salmonella enterica subsp. enterica serovar Agona (ATCC® BAA-707™)

Contig	Start	End	Name	Protein Product		EC Number	Type	RefSeq ID	Jump
1	4108772	4108350		fosfomycin resistance glutathione transferase FosA7.2	AMR		CDS	WP_000941934.1	
2	4866150	4864951	*tet(A)*	tetracycline efflux MFS transporter Tet(A)	AMR		CDS	WP_000804064.1	
2	4873227	4873700	*dfrA7*	trimethoprim-resistant dihydrofolate reductase DfrA7	AMR		CDS	WP_001389366.1	
2	4874271	4875110	*sul1*	sulfonamide-resistant dihydropteroate synthase Sul1	AMR		CDS	WP_000259031.1	

Figure 2: The AGP provides an easy-to-use Genome Browser for strains in the database.

What Reference Sequence To Use?



Personal Communications 1947: Wheeler & Stuart
Published 1948: Schaub & Hauber
1965: Elizabeth O. King
1966: Hugh & Reese
ATCC
ATCC Genome Portal (USA) ATCC 19606 (D)
AGP: 2019 — *9.5 kb pMAC*, *7.6 kb p1ATCC19606*, *52 kb Φ19606*
RM Hall (Australia)
Robert Koch Institute (Germany)
LGC (European distributor)
B. Averhoff (Germany)
~1980s
DSMZ (Germany)
~1980s
P.A.D. Grimont (France)
?
?
2014
Y. Zhu (Australia) ATCC 19606 (M) — NCBI: 2019 — *9.5 kb pMAC*, *52 kb Φ19606*
Y. Kaneko (Japan) ATCC 19606 (O) — NCBI: 2022 — *9.4 kb pMAC*, *7.6 kb p1ATCC19606*, *52 kb Φ19606*
G.S.A. Myers (Australia) ATCC 19606 (H) — NCBI: 2020 — *9.5 kb pMAC*, *7.6 kb p1ATCC19606*, *52 kb Φ19606*
P. Visca (Italy) ATCC 19606 (A) — NCBI: 2021 — *9.5 kb pMAC*, *7.6 kb p1ATCC19606*
K.J. Towner (UK) ATCC 19606 (T) — 2010 — *9.5 kb pMAC*, *7.6 kb p1ATCC19606*
H. Seifert (Germany) ATCC 19606 (S) — 2019 — *9.5 kb pMAC*, *7.6 kb p1ATCC19606*

Figure 3: Multiple assemblies of *Acinetobacter baumannii* strain 2208.

Artuso *et al.* 2022 investigated the genomes assemblies of the type strain of *Acinetobacter baumannii* in NCBI (ATCC 19606). RefSeq genomes varied by thousands of SNPs and indels with missing plasmids and prophages.

Comparisons of AGP vs NCBI RefSeq Assemblies

Yarmosh *et al.* 2022 investigated direct comparisons between ATCC genome assemblies that are in NCBI RefSeq and the AGP. Over 98% of AGP assemblies were more complete and of higher quality than RefSeq representatives.

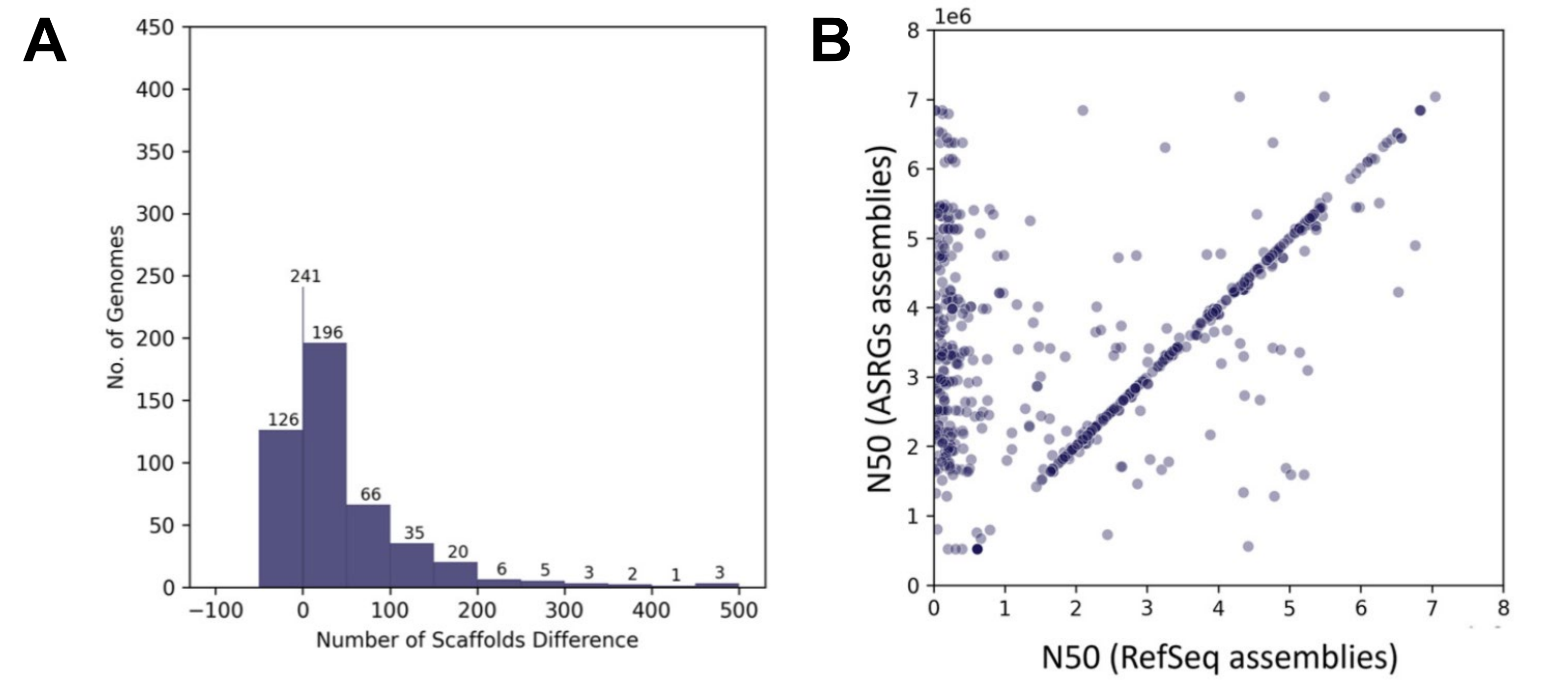


Figure 4: Comparison of AGP and NCBI RefSeq Assemblies. (A) NCBI RefSeq assemblies had more scaffolds when compared to ATCC Standard Reference Genomes (ASRGs). (B) Most ASRGs have better N50s when compared to RefSeq counterparts.

Conclusions

A majority of assemblies in AGP are of higher quality than equivalent assemblies in public genomic databases. Additionally, some of these assemblies represent strains and/or species that have no representation in public genomic databases. The AGP is a unique and authenticated resource for reference genomes.

References

1. Benton B, et al. The ATCC Genome Portal: Microbial Genome Reference Standards with Data Provenance. Microbiol Resour Announc 10(47): e0081821, 2021. PubMed: 34817215
2. Yarmosh et al. Comparative Analysis and Data Provenance for 1,113 Bacterial Genome Assemblies. mSphere 7(3): e0007722, 2022. PubMed: 35491842
3. Nguyen et al. The ATCC genome portal: 3,938 authenticated microbial reference genomes. Microbiol Resour Announc 13(2): e0104523, 2024. PubMed: 38289057
4. Nguyen et al. Reclassification of atypical *Moraxella catarrhalis* ATCC 23246 as *Moraxella veridica* sp. nov. Int J Syst Evol Microbiol 75(5): ePub. PubMed: 40397495
5. Artuso I, Lucidi M, Visaggio D, Capecchi G, Lugli GA, Ventura M, Visca P. Genome diversity of domesticated Acinetobacter baumannii ATCC 19606T strains. Microb Genom. 2022 Jan;8(1):000749. PubMed: 8914354.

Learn more about the ATCC® Genome Portal

Visit and explore the ATCC® Genome Portal