EEB-SUNDAY-1139
June 22, 2025

Mining the ATCC Genome Portal: Genomic Insights into Unsequenced Taxa and Rare Historical Strains in the American Type Culture Collection

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

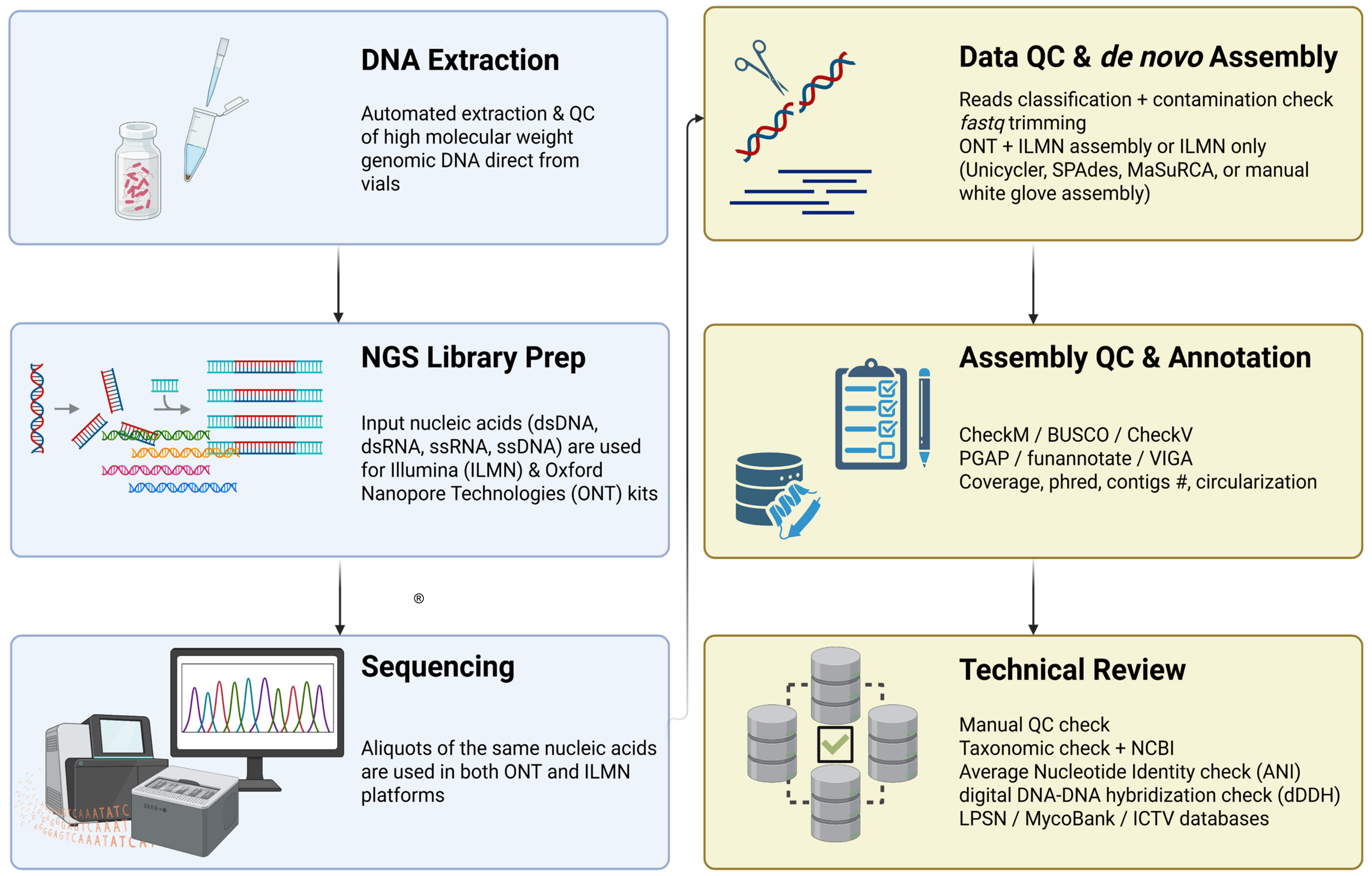
Introduction

The ATCC® Genome Portal (AGP) serves as the “digital twin” of a 100-year-old culture collection. Many of the authenticated reference genomes in the AGP represent strains commonly used in research and as standards.

Strains deposited into ATCC® over the past century represent the collective efforts of the research community to preserve cultures they deemed important. Many of these strains have never been sequenced and are understudied.

The AGP initiative presents an opportunity to not only provide an authenticated reference genome of strains in the collection, but to also reassess the taxonomic identity of these strains. Additionally, these genomes can provide insight into novel biosynthetic gene clusters. This work presents some insights into some of these understudied strains.

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

Mining for Potential New Bioactive Natural Products

ATCC® 31571™ was originally deposited as a *Pseudomonas* sp. It produces Bactobolin A with reported broad antibiotic and antitumor activities (Figure 2). Next-generation sequencing (NGS) shows that ATCC® 31571™ is an undescribed novel *Burkholderia* sp.

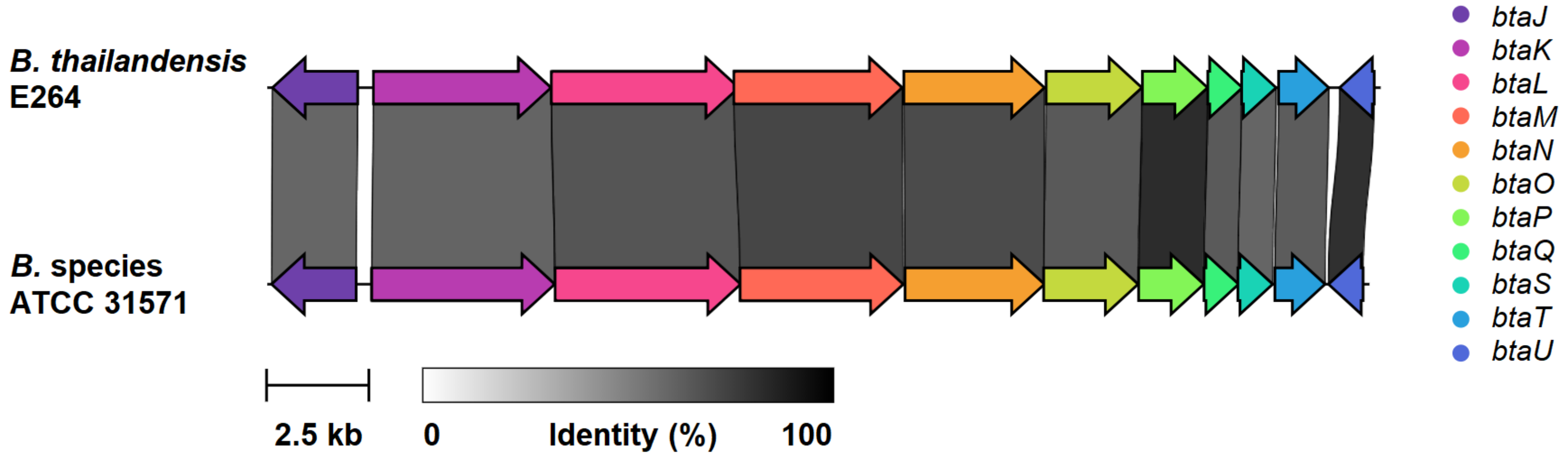


Figure 2: Bactobolin A biosynthetic gene clusters (BGCs) between *Burkholderia thailandensis* and ATCC® 31571™.

ATCC® 36554™ was originally deposited as a mycorrhizal fungus and after NGS, was identified as a *Lachnum* species. BCGs identified by antiSMASH show interesting natural product candidates (Figure 3).

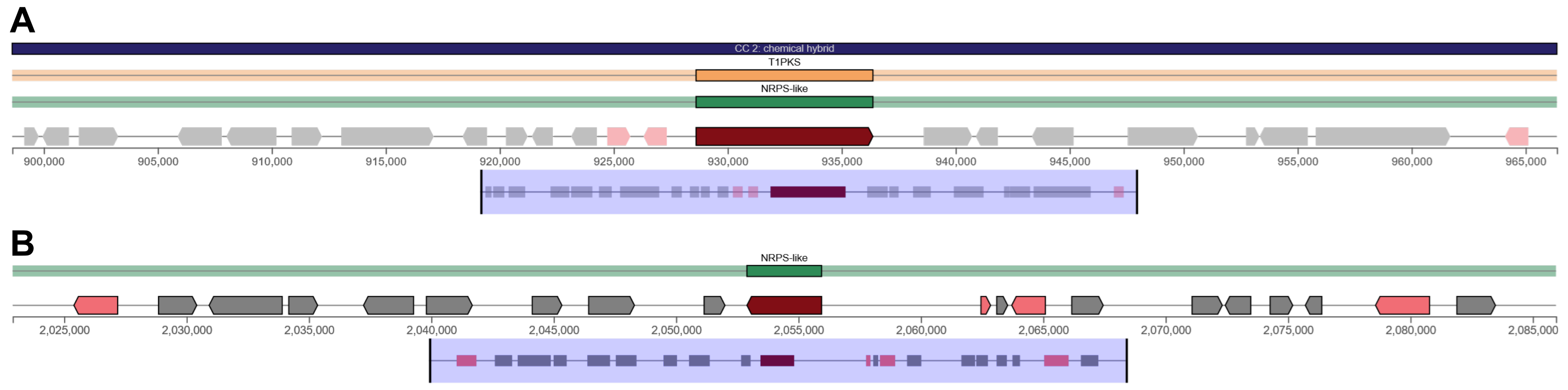


Figure 3: BGCs identified in ATCC® 36554™. (A) A swainsonine-like cluster (antitumor). (B) An aspulvinone-like cluster (antiviral).

Discovering Reference Genomes for Unsequenced Taxa

Table 1: Examples of unsequenced taxa in the AGP.

ATCC® No.	Species	AGP Link*
700426™	*Desulfocella halophila*	https://genomes.atcc.org/genomes/69b18767b1624583
BAA-1173™	*Methanomethylovorans thermophila*	https://genomes.atcc.org/genomes/b57ea71419f645f9
700261™	*Spirochaeta asiatica*	https://genomes.atcc.org/genomes/5cb3037d833d4a3a
BAA-262™	*Gelria glutamica*	https://genomes.atcc.org/genomes/15bd28ab7da14589
35158™	*Filomicrobium fusiforme*	https://genomes.atcc.org/genomes/d9b15c5dd55c4b12
700687™	*Zoogloea resiniphila*	https://genomes.atcc.org/genomes/d6fd5f445ec14336

*Nearly 300 genomes in the AGP are not represented by sequences in NCBI.

Revising Taxonomic Classifications

The AGP also provides opportunities to revise taxonomic classification of historical strains in the collection. Recently, the identity of ATCC® 23246™ is now resolved by NGS after decades of misclassification as *Moraxella catarrhalis* (Figure 4).

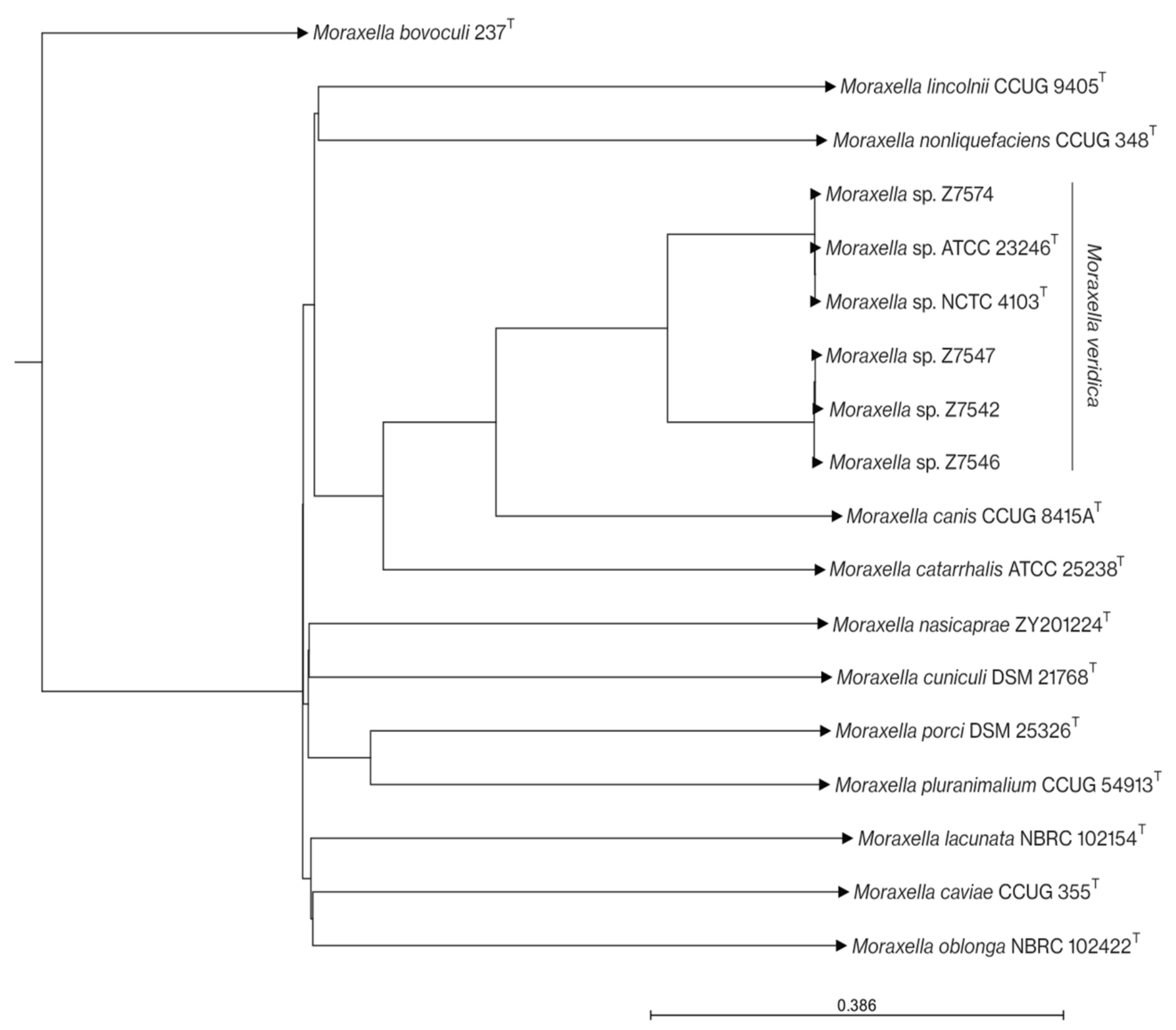


Figure 4: Phylogenetic tree of a commonly used and misclassified *M. catarrhalis* strain. This work resulted in the description of a new species in the *Moraxella* genus: *Moraxella veridica*.

Conclusions

By sequencing understudied and rare strains in the collection, the AGP initiative will provide the research community important insights into a vast repository of microbes that have been collected in the past century. These genomic sequences will help uncover novel natural products and new species.

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

Learn more about the ATCC® Genome Portal

Visit and explore the ATCC® Genome Portal

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