

High-quality genome assemblies and biosynthetic gene clusters annotation from laboratory reference fungal strains

Ford Combs, MS; Juan Lopera, PhD; Nikhita Puthuveetil, MS; David Yarmosh, MS; Samuel Greenfield, BS; Stephen King, MS; Marco Riojas, PhD; Amanda Pierola, BS; John Bagnoli, BS; Briana Benton, BS; Jonathan Jacobs, PhD
ATCC, Manassas, VA 20110



Background

Fungi produce a wide variety of secondary metabolites that play an important role in human health and disease and have great potential as drug candidates and therapeutics. A comparative analysis of public assemblies and ATCC's internally produced assemblies showed that ATCC's assemblies typically had higher quality. Further analysis using genome annotation and mining tools indicated high levels of biosynthetic potential in the ATCC mycology collection.

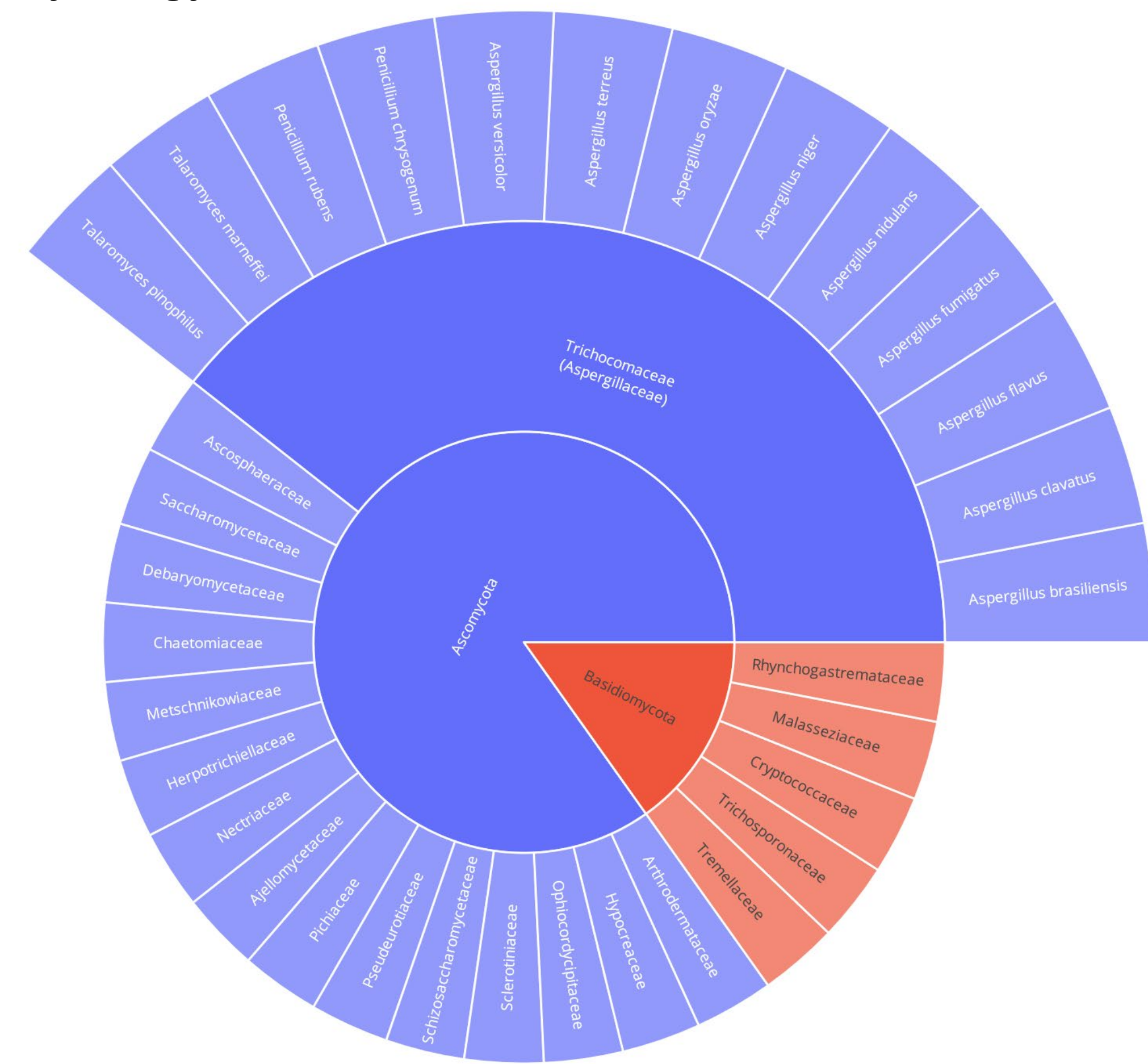


Figure 1: Taxonomic diversity of fungi on the ATCC Genome Portal. The innermost partition includes two phyla: *Ascomycota* and *Basidiomycota*; the middle partition includes each of the families found in the collection; and the outermost partition displays each of the species of the *Trichocomaceae* family found in the collection. While many biosynthetic gene clusters (BGCs) were found in the mycology collection, a particularly large and diverse set was found in *Trichocomaceae*.

ATCC Enhanced Authentication Initiative: Mycology Collection

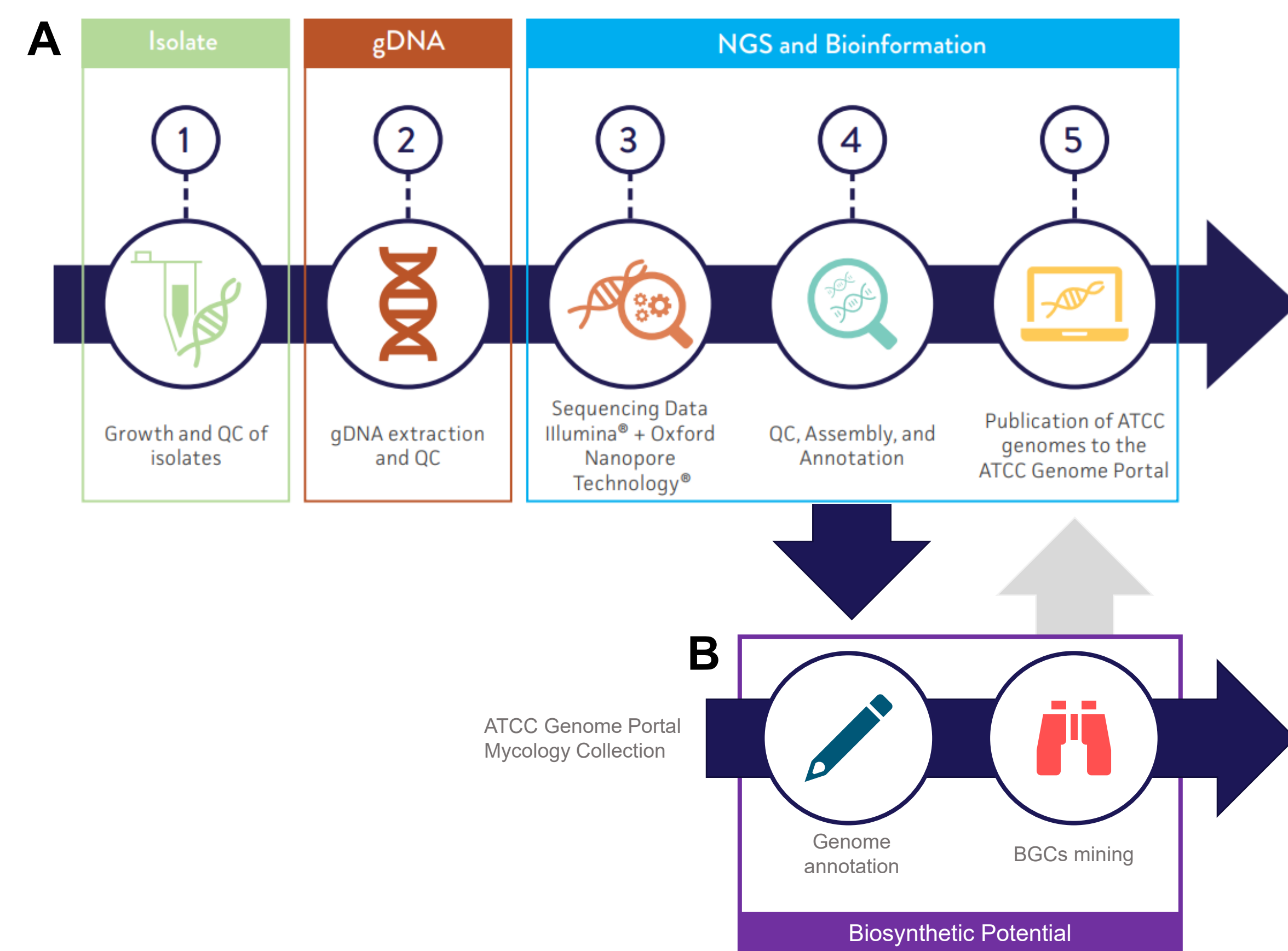


Figure 2: Workflows related to this study. (A) The ATCC genome portal workflow used to produce high-quality, authenticated genomes. (B) The biosynthetic potential workflow used in this study. ATCC Mycology Collection genomes were annotated using BRAKER and then mined using antiSMASH to detect BGCs.

Comparison of ATCC and Public Assembly Quality

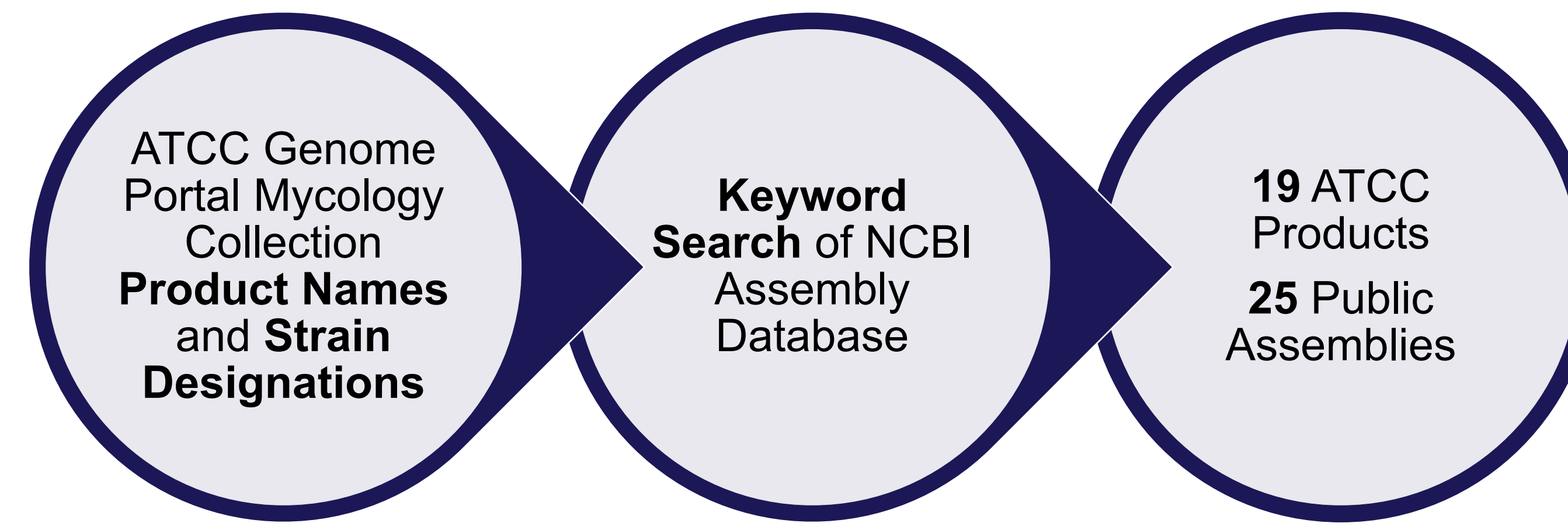


Figure 3: Process description for finding public assemblies associated with ATCC mycology collection products. The product name and strain designations for each ATCC mycology collection product were used as keywords in a search of the NCBI assembly database.

Comparison of ATCC and NCBI Assembly Quality

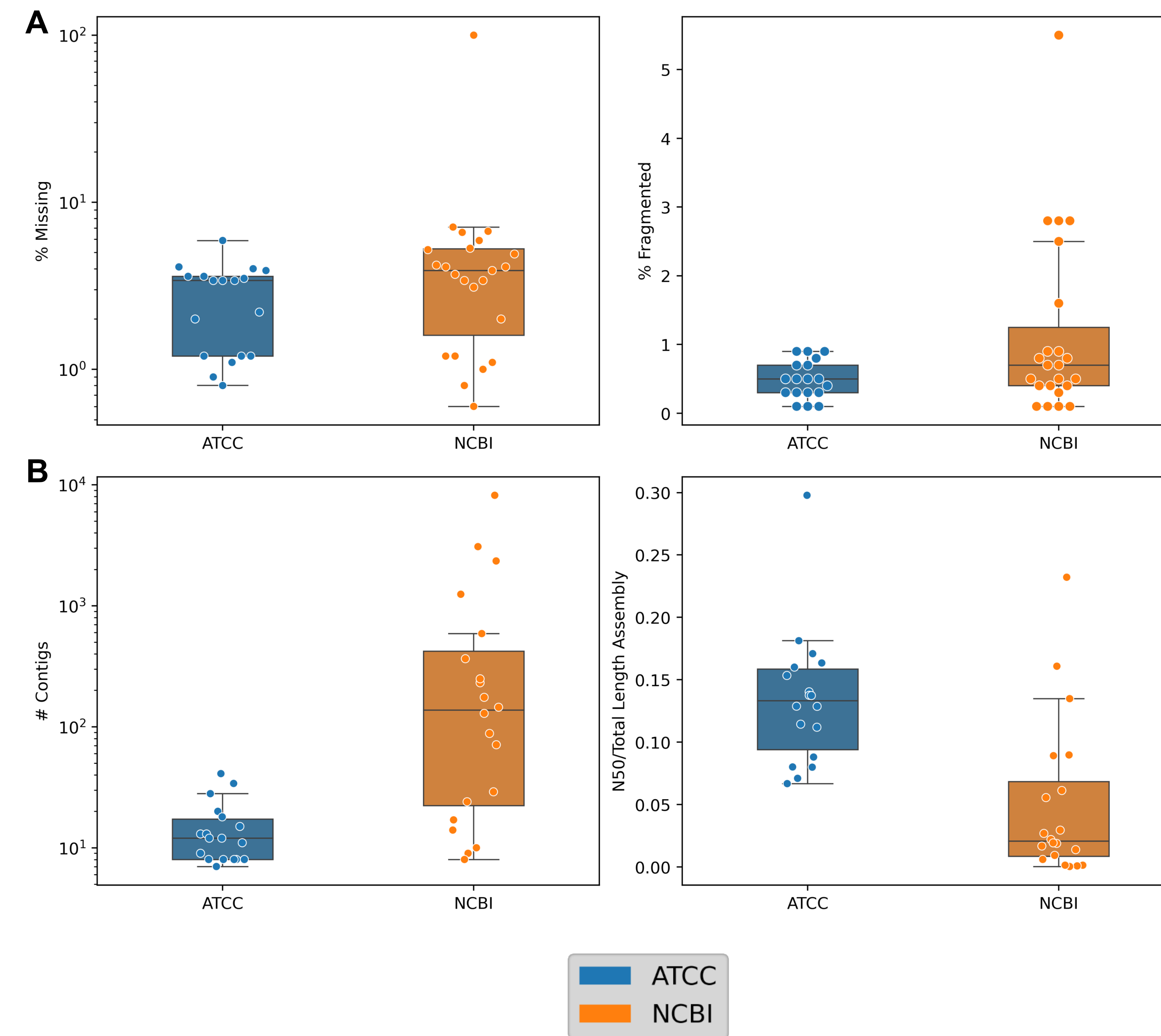


Figure 4: Comparison results of ATCC Genome Portal assemblies and their associated public assemblies. (A) Scatterplots showing BUSCO percent missing and percent fragmented of each assembly. (B) Scatterplots showing contig count and N50/Total Length of Assembly of each assembly.

BGC Mining and Phylogenetic Analysis

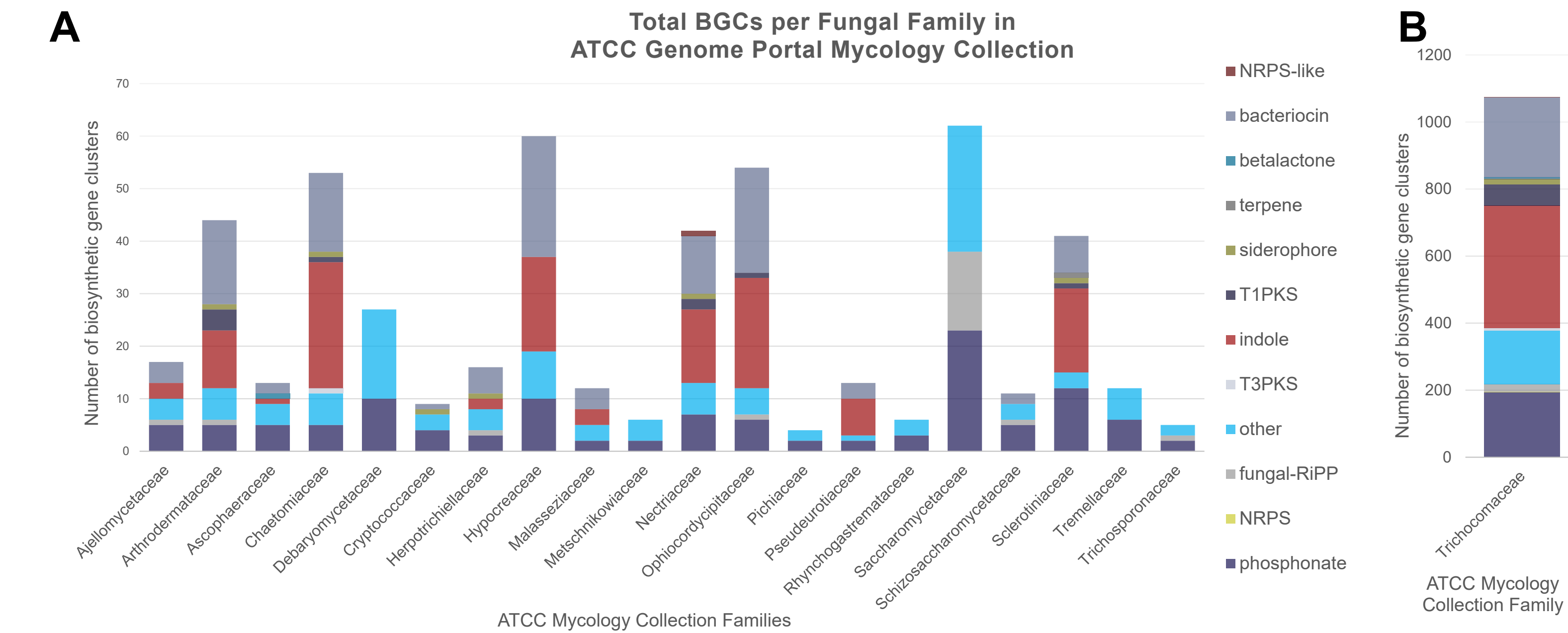


Figure 5: Charts depicting the total number of BGCs of different types found in the fungal families from the ATCC Genome Portal Mycology Collection. (A) All fungal families except for *Trichocomaceae* and (B) *Trichocomaceae* only, separated because of its significantly greater quantity of BGCs.

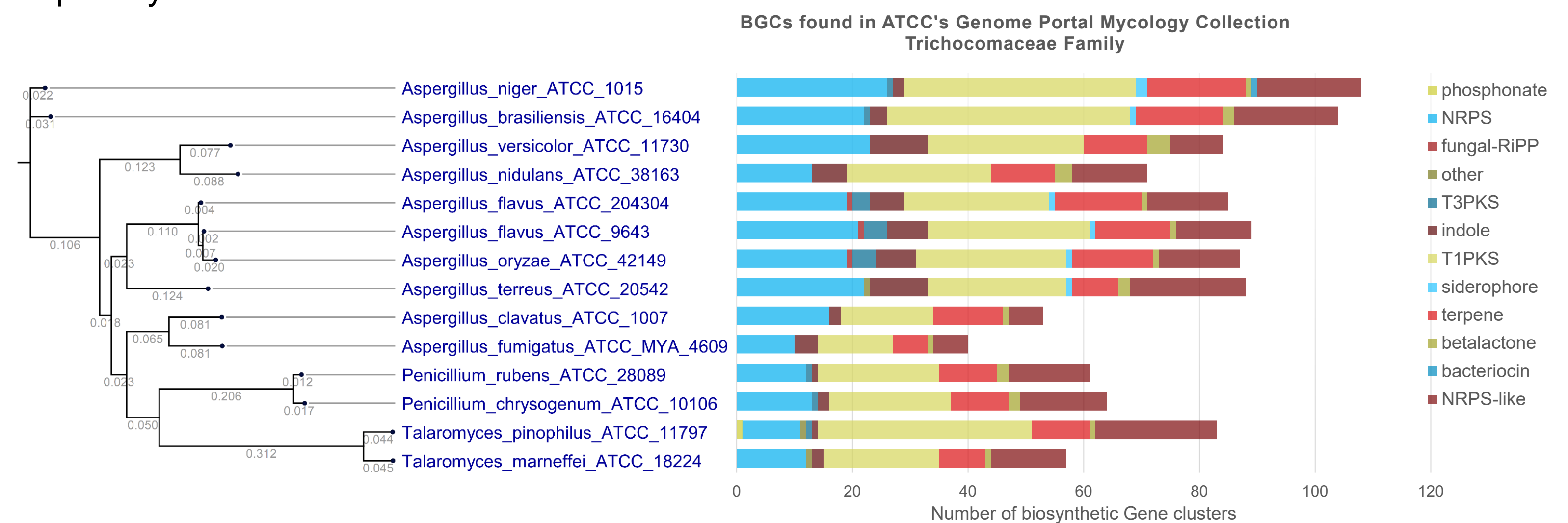


Figure 6: Phylogenetic tree (ML) of *Trichocomaceae* family and associated number and types of BGCs found in each species.

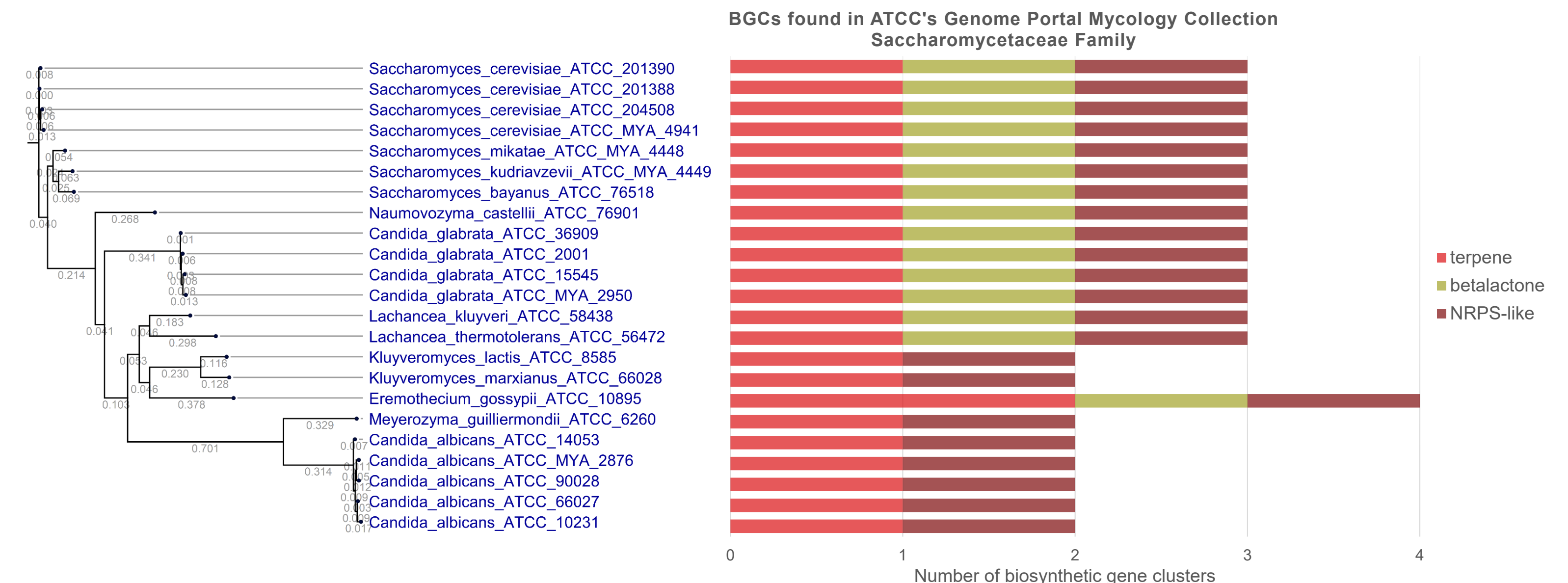


Figure 7: Phylogenetic tree (ML) of *Saccharomycetaceae* family and associated number and types of BGCs found in each species.

Conclusions

The ATCC Genome Portal mycology collection contains dozens of high-quality, authenticated assemblies that are linked to physical materials. Annotation and genome mining revealed a large presence of BGCs in these genomes, particularly in the *Trichocomaceae* family.