#3960
June 12, 2022

# Increasing Oxford Nanopore Technologies Throughput for the ATCC Genome Portal

## Expanding Genome Publication Rates through an Improved Understanding of Bacterial Genome Assemblies

ATCC®
Credible leads to Incredible®

Samuel R. Greenfield, BS; Nikhita Puthuveetil, MS; James Duncan, BS; Amanda Pierola, BS; Corina Tabron, MS; David Yarmosh, MS; Amy Reese, MS; Stephen King, MS; Robert Marlow, BS; Ford Combs, PhD; Ana Fernandes, BS; John Bagnoli, BS; Briana Benton, BS; Jonathan Jacobs PhD
ATCC, Manassas, VA 20110

## Background

ATCC has made a commitment to the scientific community to provide reference-quality, whole-genome sequencing data for items within our collection through the ATCC Genome Portal. Our Enhanced Authentication Initiative enriches the characterization of our biological collections by using next-generation sequencing to provide reference-quality genomes to the scientific community; however, given the fact that we have only published 2,158 of our 78,700 items, a higher throughput method must be implemented to increase our publication rate. We sought to evaluate our current DNA requirements to determine if lower thresholds could generate similar quality bacterial assemblies, allowing for more rapid and automated DNA extractions and therefore an increased number of genomes being published to the ATCC Genome Portal.

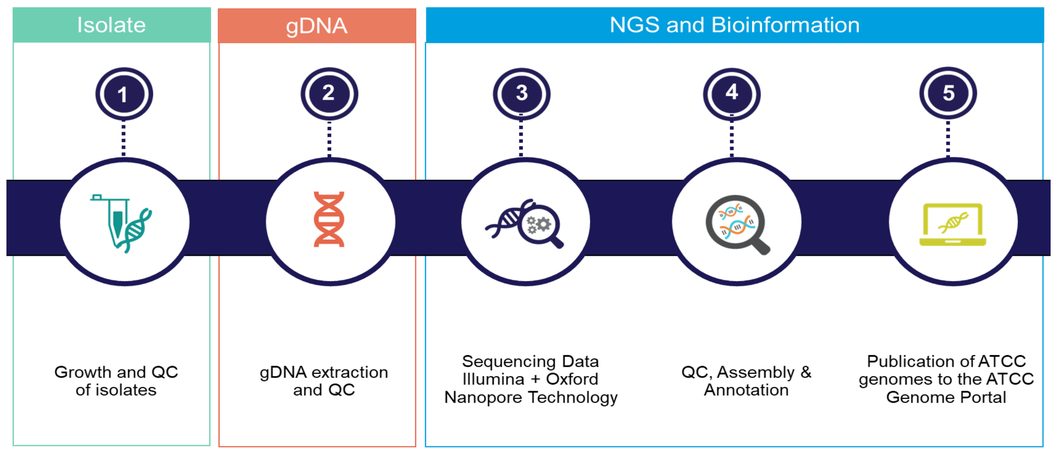



**Figure 1.** A hybrid assembly method combining Illumina® and Oxford Nanopore Technologies® (ONT) data to assemble the highest quality bacterial genomes is used.

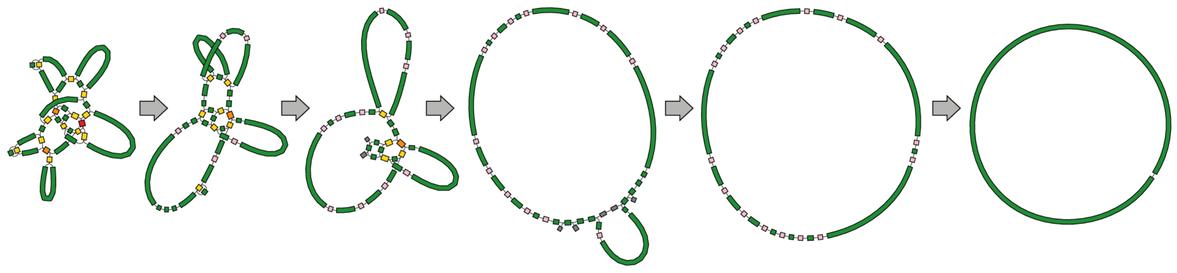
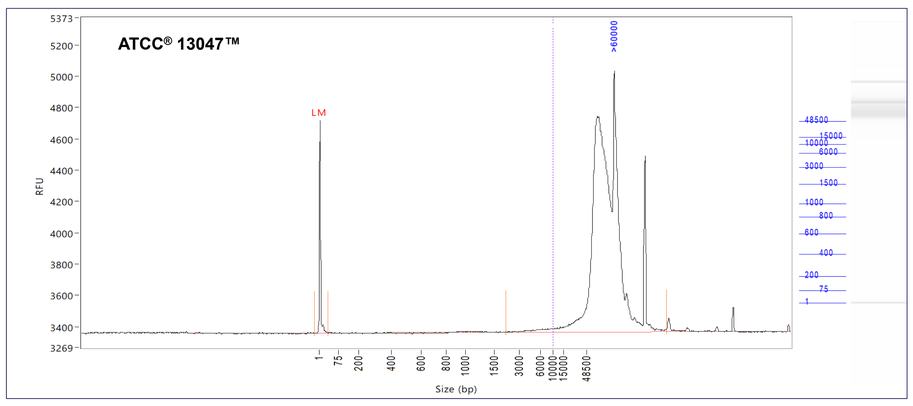


**Figure 2.** ONT long-reads are used to resolve tangled Illumina-only assemblies. To be published to the ATCC Genome Portal, bacterial assemblies are required to have <30 contigs, >100X Illumina depth, and both >95% completeness and <5% contamination analyzed by checkM.[2] Image reproduced from https://github.com/rrwick/Unicycler.




| ATCC® | Organism | Qubit® (ng/µL) | $A_{260}/A_{280}$ Ratio | Mean fragment size (kb) |
|---|---|---|---|---|
| 13047™ | *Enterobacter cloacae* subsp. *cloacae* | 32.7 | 1.85 | >60 |

**Figure 3.** We currently require DNA fragments for ONT sequencing to have an average fragment length of 20 kb, 50% of fragments to be larger than 10 kb, and at least 1000 ng of starting material. These standards are only consistently achieved using QIAGEN® Genomic-Tip (catalog no. 10223, QIAGEN, MD, USA), a time-consuming kit but in line with current ONT standards for the native barcoding and ligation sequencing with a R9.4.1 flow cell.

## Methods

A hybrid assembly was performed using Unicycler[3] with default parameters for each read set (see table 1) using filtered ONT and unfiltered Illumina reads. This constituted 63 separate assemblies. All ONT data was generated by native barcoding gDNA using LSK109 and NBD-104/114 on an R9.4.1 flow cell (Oxford Nanopore Technologies, UK). Illumina short read data was generated on a MiSeq™ from libraries created with Illumina's DNA Prep kit and unique dual indexes (catalog no. 20018705 and 20027213, Illumina, CA, USA).

**Table 1.** Illumina and Nanopore fastq files for nine bacterial samples published to the ATCC Genome Portal were gathered, evaluated using NanoPlot,[1] and filtered to maximum length thresholds of 5, 7.5, 10, 15, 20, 25-200 kb, and no threshold using Filtlong.[4] Cells highlighted red failed to meet our previously stated publication standards while cells highlighted green met these standards.

| ATCC® | Genome Size (bp) | 5kb | | 7.5kb | | 10kb | | 15kb | | 20kb | | 25-200kb | | No Max Filter | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | ONT Reads | N50 | ONT Reads | N50 | ONT Reads | N50 | ONT Reads | N50 | ONT Reads | N50 | ONT Reads | N50 | ONT Reads | N50 |
| 23114™ | 669,676 | 6,341 | 4,348 | 4,159 | 6,642 | 3,122 | 8,836 | 2,105 | 13,171 | 1,586 | 17,519 | 554 | 50,961 | 127,862 | 25,386 |
| 33453™ | 793,473 | 8,173 | 4,110 | 5,236 | 6,352 | 3,869 | 8,550 | 2,565 | 12,973 | 1,976 | 17,007 | 786 | 41,796 | 101,107 | 24,675 |
| 49145™ | 1,716,585 | 9,635 | 3,398 | 13,630 | 4,848 | 11,605 | 6,869 | 6,592 | 11,333 | 4,912 | 15,124 | 1,668 | 42,433 | 33,015 | 24,708 |
| 14799™ | 2,207,181 | 33,589 | 3,092 | 20,362 | 4,740 | 15,362 | 6,265 | 10,209 | 9,599 | 7,665 | 12,783 | 2,032 | 45,144 | 61,016 | 25,593 |
| 33236™ | 2,281,739 | 4,575 | 3,340 | 6,336 | 4,757 | 7,581 | 6,013 | 9,261 | 8,398 | 10,341 | 10,519 | 2,435 | 38,019 | 13,756 | 25,088 |
| 13047™ | 5,608,020 | 15,128 | 3,202 | 19,433 | 4,415 | 22,629 | 5,680 | 26,817 | 8,059 | 29,732 | 10,376 | 5,527 | 41,574 | 37,777 | 25,076 |
| 700441™ | 6,570,388 | 83,574 | 3,356 | 55,942 | 4,981 | 43,089 | 6,515 | 29,975 | 9,496 | 23,048 | 12,568 | 5,418 | 51,859 | 239,555 | 25,250 |
| 700084™ | 6,947,810 | 21,809 | 3,232 | 29,470 | 4,662 | 36,047 | 6,283 | 31,033 | 10,627 | 18,933 | 15,799 | 7,821 | 35,689 | 91,493 | 25,225 |
| BAA-477™ | 7,070,224 | 82,821 | 2,955 | 78,957 | 4,215 | 49,360 | 6,352 | 29,131 | 10,465 | 21,281 | 14,370 | 6,274 | 47,351 | 158,833 | 25,417 |

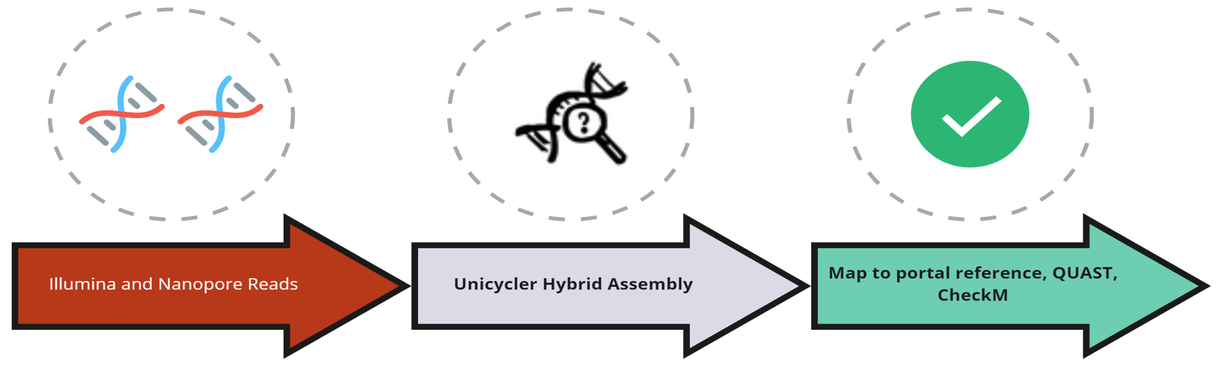



**Figure 4.** The resulting assemblies were evaluated by mapping assemblies to our own ATCC Genome Portal reference, running CheckM[2] for completeness and contamination metrics and QUAST[5] for assembly metadata. Each was evaluated based on current ATCC Genome Portal publication requirements.

## Results

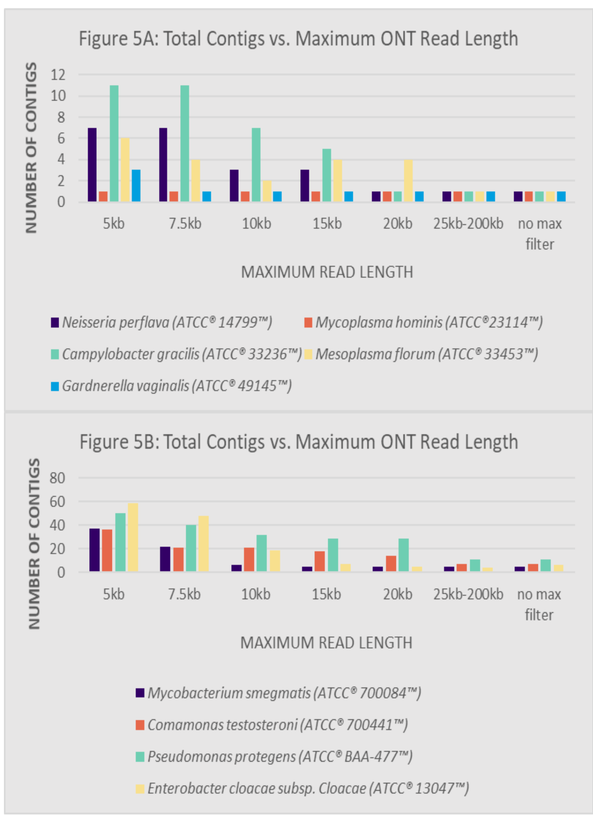



**Figure 5(A&B).** Read length with corresponding number of contigs for each bacterial genome assembly.

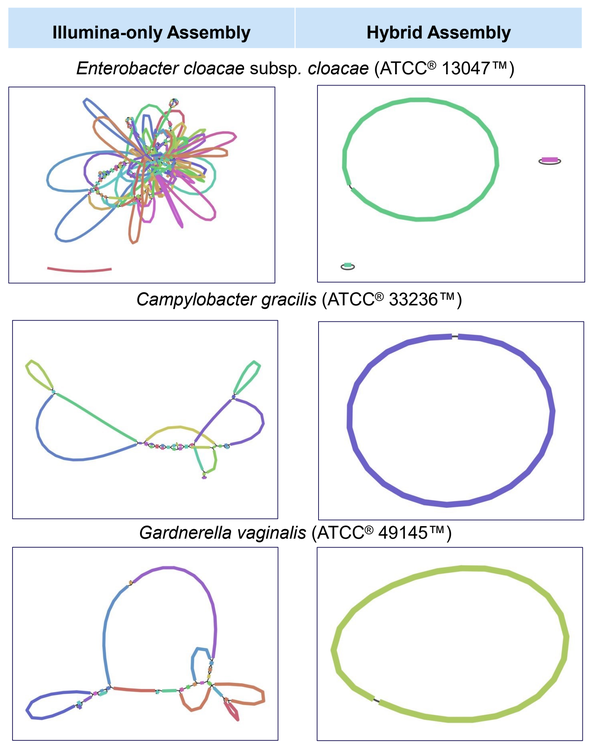



**Figure 6.** Visualization of Illumina-only and hybrid de novo assemblies. From our analysis, at ONT read lengths >8-12 kb, the hybrid assemblies meet ATCC's requirements for publication.

Our study provides in silico evidence to credit hybrid assemblies derived from smaller fragments. As seen by the filtered maximum 15 kb read N50s, we are confident that an average fragment length of >8-12 kb will provide successful high quality bacterial genome assemblies for ATCC's Genome Portal. With the decrease in average fragment length, we have been able to implement QIAGEN EZ1® extraction robots (QIAGEN, MD, USA), which have led to a 4X increase in extraction efficiency compared to the QIAGEN genomic-tip.

**References**
1. De Coster, Wouter (2016) NanoPlot [Source code] https://github.com/wdecoster/NanoPlot
2. Parks, D, Skennerton, C, Imelfort, M (2014) CheckM [Source code] https://github.com/Ecogenomics/CheckM
3. Wick, R (2016) Unicycler [Source code] https://github.com/rrwick/Unicycler
4. Wick, R (2017) Filtlong [Source code] https://github.com/rrwick/Filtlong
5. Gurevich, A, et al. (2018) QUAST [Source code] https://github.com/ablab/quast

Learn more about our Enhanced Authentication Initiative

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