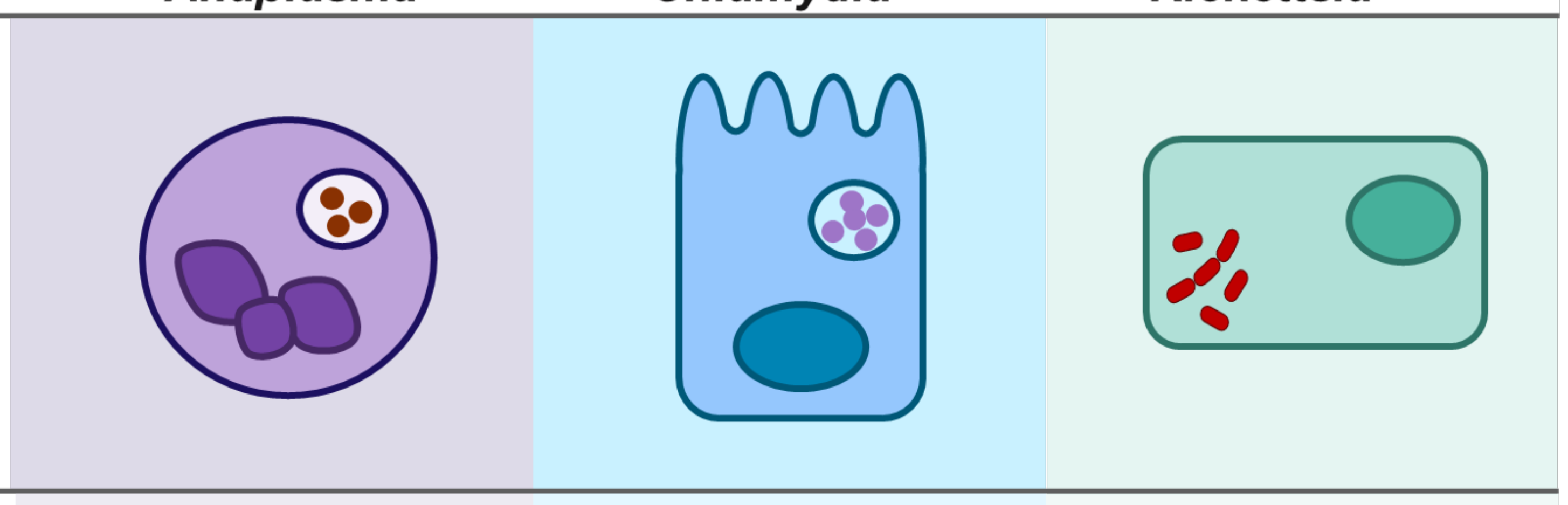
Inside Job: Targeted Assembly of Obligate Intracellular Bacterial Genomes

Nikhita Puthuveetil, MS; Joseph Petrone, PhD; Amy Reese, MS; David Yarmosh, MS; John Bagnoli, BS; and Jonathan Jacobs, PhD
ATCC, Manassas, VA 20110

Obligate Intracellular Bacteria

Obligate intracellular bacteria (OIB), unlike other bacterial genera, are entirely dependent upon their eukaryotic hosts for survival. Members of this group include species from the *Chlamydia*, *Anaplasma*, and the *Rickettsia* genera, many of which are pathogenic to humans. Due to their absolute dependence on their hosts, genome sequencing and assembly of OIB can be complicated by the inability to separate bacterial sequences from host DNA. Here, we present ATCC®'s OIB assembly pipeline.



	Anaplasma	*Chlamydia*	*Rickettsia*
Genome Size	1-1.5 Mb	1-1.2 Mb	1.1-2 Mb
GC Percentage	41.5-49%	37.5-42%	29-32.5%
Transmission	vector	direct contact	vector
Transmission mode	arthropod-human	human-human	arthropod-human
Host cell tropism	neutrophils & erythrocytes	epithelial	endothelial
Replication niche	vacuole	vacuole	cytoplasm
Associated Diseases	Anaplasmosis	Chlamydia	Rocky Mountain Spotted Fever

Figure 1: Characteristics of OIB. Genomes of OIB are often quite small as compared to other bacteria, and their transmission mode and host cell tropism can vary widely across genera.

OIB Pipeline Design

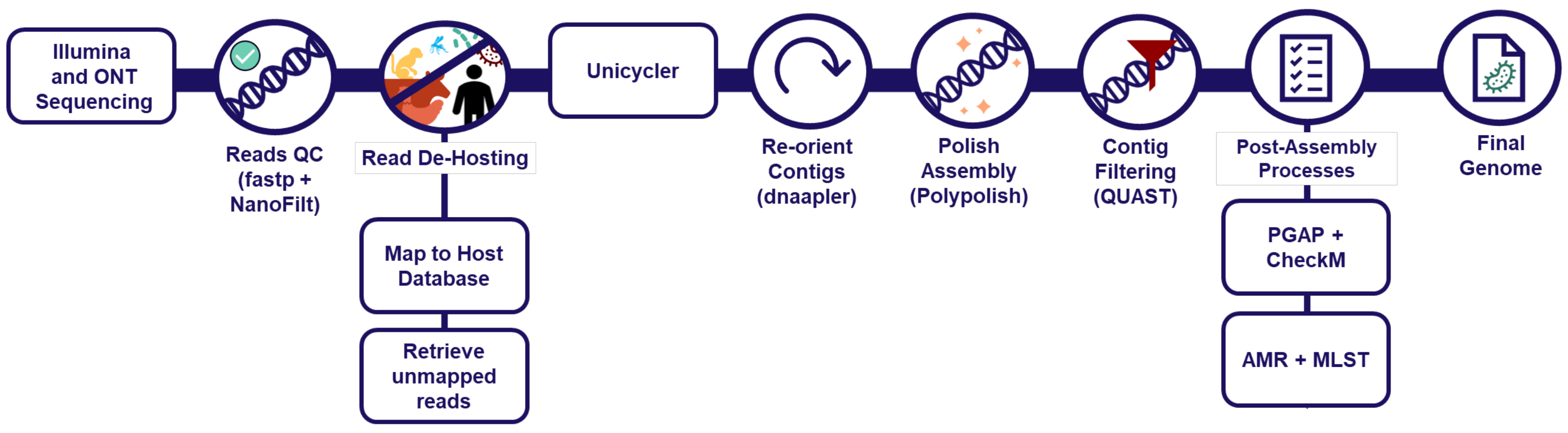


Figure 2: OIB pipeline. OIB go through ATCC®'s standard bacterial assembly pipeline with the addition of a read de-hosting step which involves mapping both short and long reads to a custom host database. Reads that do not map to the database are considered "non-host" and are used as input through the rest of the pipeline.

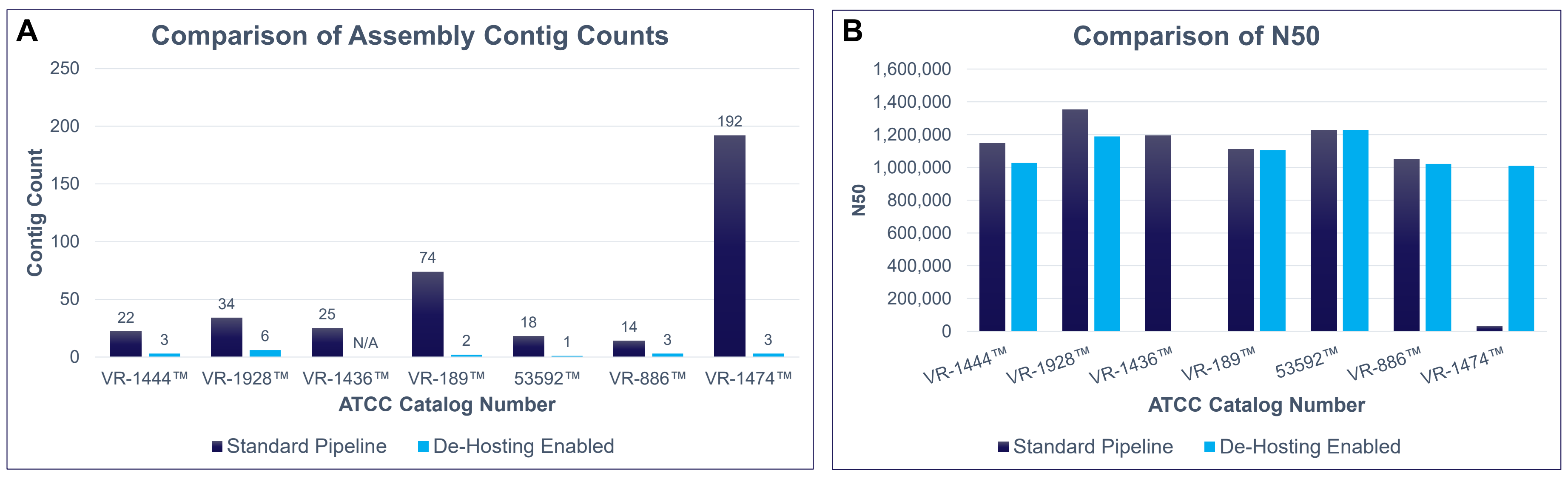
Table 1: Test bacteria. Twelve different OIB from a variety of genera were used to test the performance of the ATCC® OIB pipeline.

OIB Name	Additional Notes	Available Genomes in NCBI
Rickettsia canadensis	Unclear pathogenicity, infection observed in humans	2
Rickettsia tamurae subsp. *buchneri*	Tick symbiont	2
Rickettsia monacensis	Causes acute tickborne rickettsiosis in humans	1
Anaplasma marginale	Causes acute infection in cattle	24
Chlamydia muridarum	Causes chlamydia in mice	35
Chlamydia trachomatis	Causes chlamydia in humans	385
Chlamydia pneumoniae	Causes pneumonia in humans	16
Chlamydia pecorum	Causes acute polyarthritis in lambs	28
Chlamydia abortus	Causes abortion in mammals	74
Chlamydia caviae	Causes conjunctivitis and respiratory infections in guinea pigs	2

Pipeline Comparisons

Table 2: Results summary. Assemblies generated with our standard bacterial pipeline have acceptable lengths when compared to the expected species length. However, when de-hosting was enabled, the generated bacterial assemblies are slightly closer in length.

			Assembly Length	
Catalog Number	**Name**	**Expected Length (NCBI)**	**Standard Pipeline**	**De-Hosting Enabled**
ATCC® VR-1444™	*Rickettsia canadensis*	1,150,228	1,344,391	1,160,365
ATCC® VR-1928™	*Rickettsia monacensis*	1,353,450	1,664,611	1,386,810
ATCC® VR-1436™	*Anaplasma marginale*	1,202,435	1,324,627	N/A
ATCC® VR-189™	*Chlamydia pecorum*	1,106,403	1,664,411	1,113,162
ATCC® 53592™	*Chlamydia pneumoniae*	1,225,935	1,371,451	1,227,376
ATCC® VR-886™	*Chlamydia trachomatis*	1,042,519	1,174,501	1,046,096
ATCC® VR-1474™	*Chlamydia suis*	1,094,002	1,508,837	1,033,562



Figures 4: Contig counts and completeness comparison. (A) The de-hosting enabled pipeline drastically reduced contig counts in OIBs compared to the standard pipeline. (B) However, the n50s for assemblies generated by each pipeline stayed about the same.

OIB Methylation Data

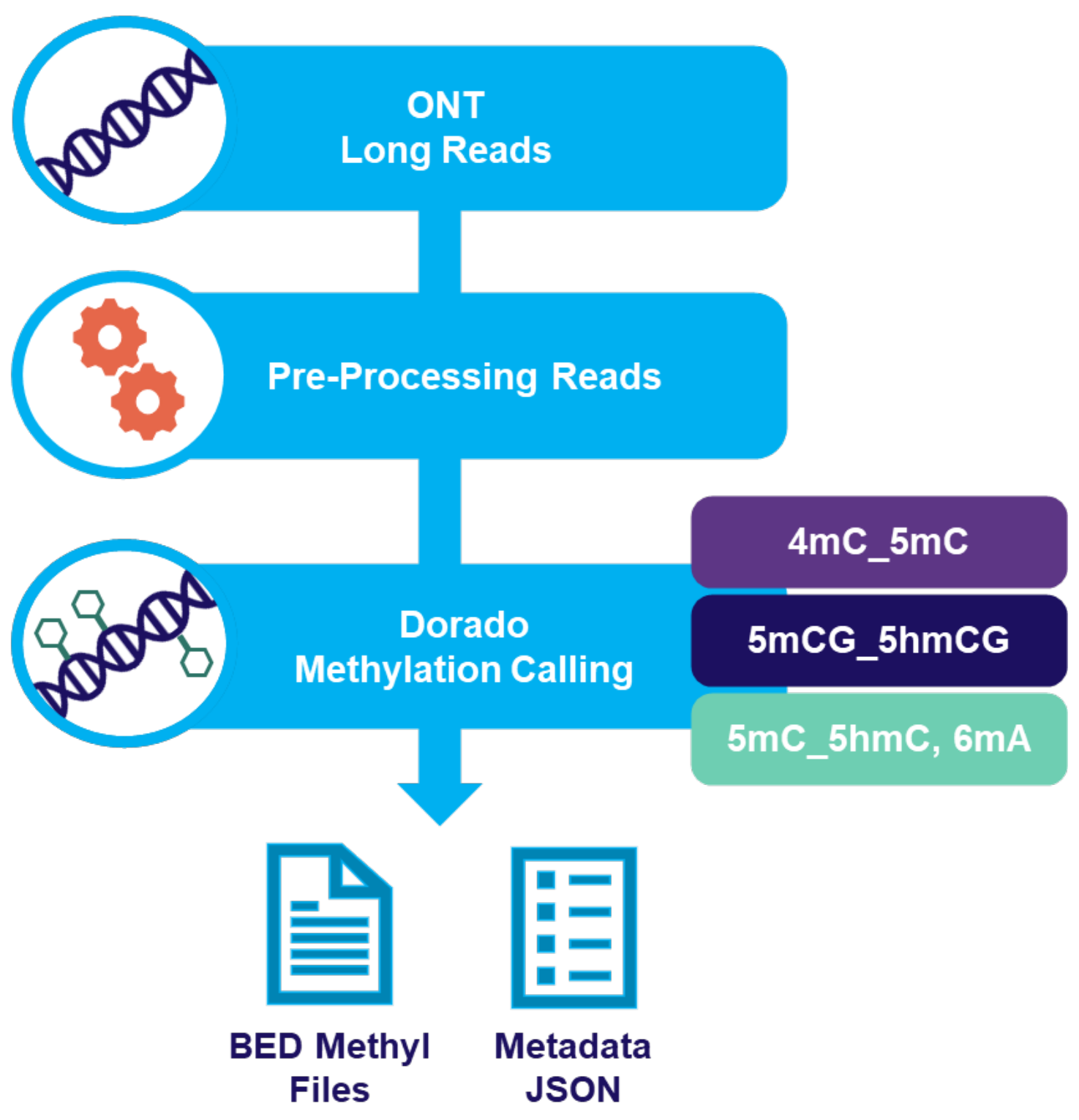
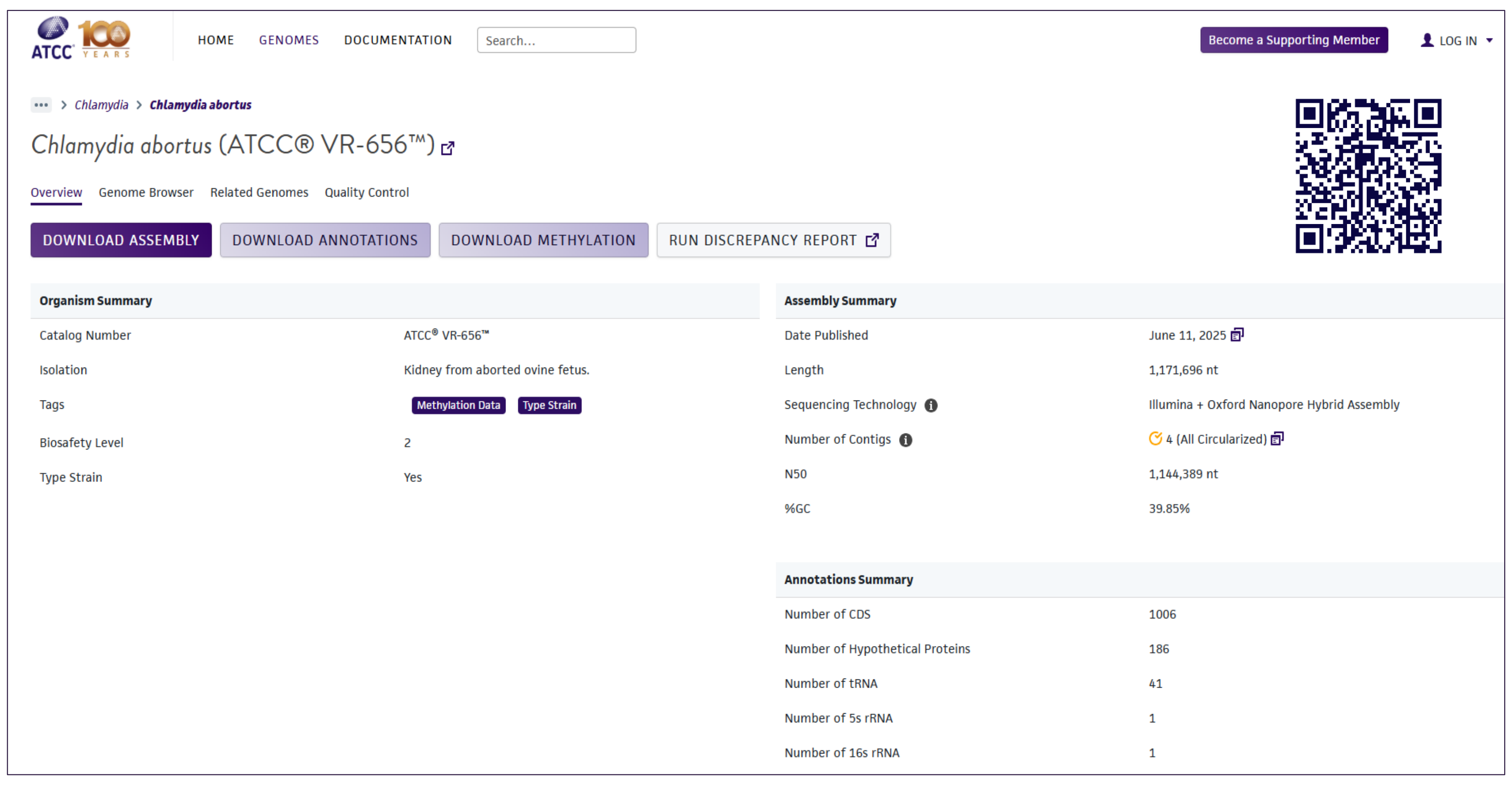


Figure 5: Methylation calling pipeline. All bacteria published to the ATCC® Genome Portal (AGP) go through long-read methylation calling. Methylation data can provide insight to the mechanisms which OIBs use to survive in host cells, which is still not quite understood. In this pipeline, several different methylation models are used.



••• > *Chlamydia* > ***Chlamydia abortus***

Chlamydia abortus (ATCC® VR-656™)

Organism Summary

Catalog Number	ATCC® VR-656™
Isolation	Kidney from aborted ovine fetus.
Tags	Methylation Data Type Strain
Biosafety Level	2
Type Strain	Yes

Assembly Summary

Date Published	June 11, 2025
Length	1,171,696 nt
Sequencing Technology	Illumina + Oxford Nanopore Hybrid Assembly
Number of Contigs	4 (All Circularized)
N50	1,144,389 nt
%GC	39.85%

Annotations Summary

Number of CDS	1006
Number of Hypothetical Proteins	186
Number of tRNA	41
Number of 5s rRNA	1
Number of 16s rRNA	1

Figure 6: OIB on the AGP. An example page of VR-656™ is shown. All OIBs that met our publication criteria as a result of this analysis were or will be published to the AGP.

Conclusion

The mechanisms that OIBs employ to survive within the host cell are not yet well understood, but through the assembly of their genomes, we can elucidate the unique host-pathogen interactions that allow OIBs to effectively replicate in their host cells. In this analysis, de-hosting reads prior to assembly led to viable, complete assemblies. Based on these results, this method will be used moving forward to publish OIBs to the AGP.

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