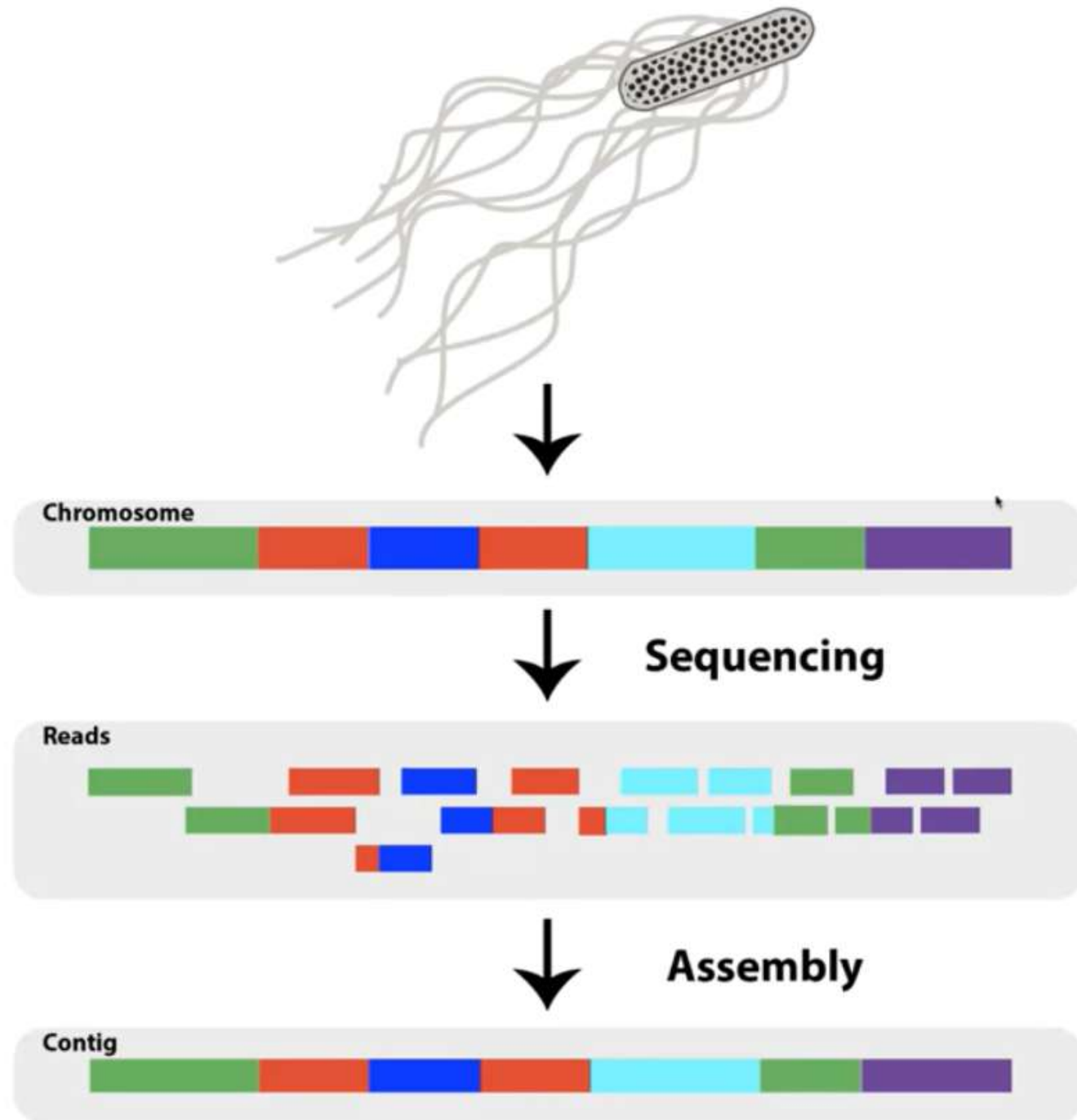


# Genome Assembly

What does it mean to assemble a genome?

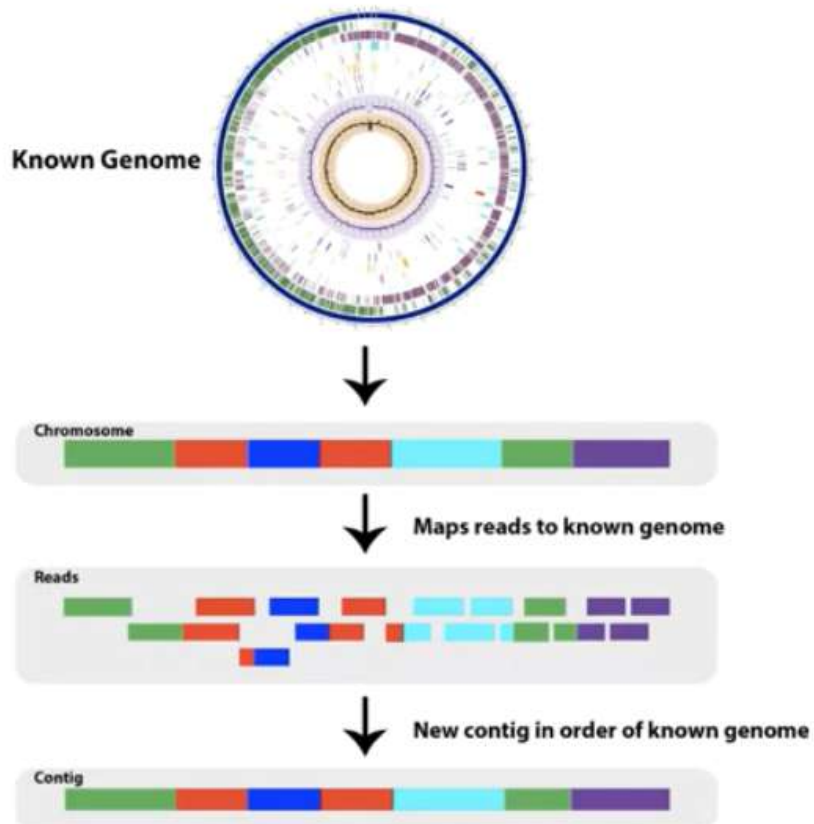
# Genome assembly

- Chromosomes fragmented
- Fragments sequenced
- Sequences put back together

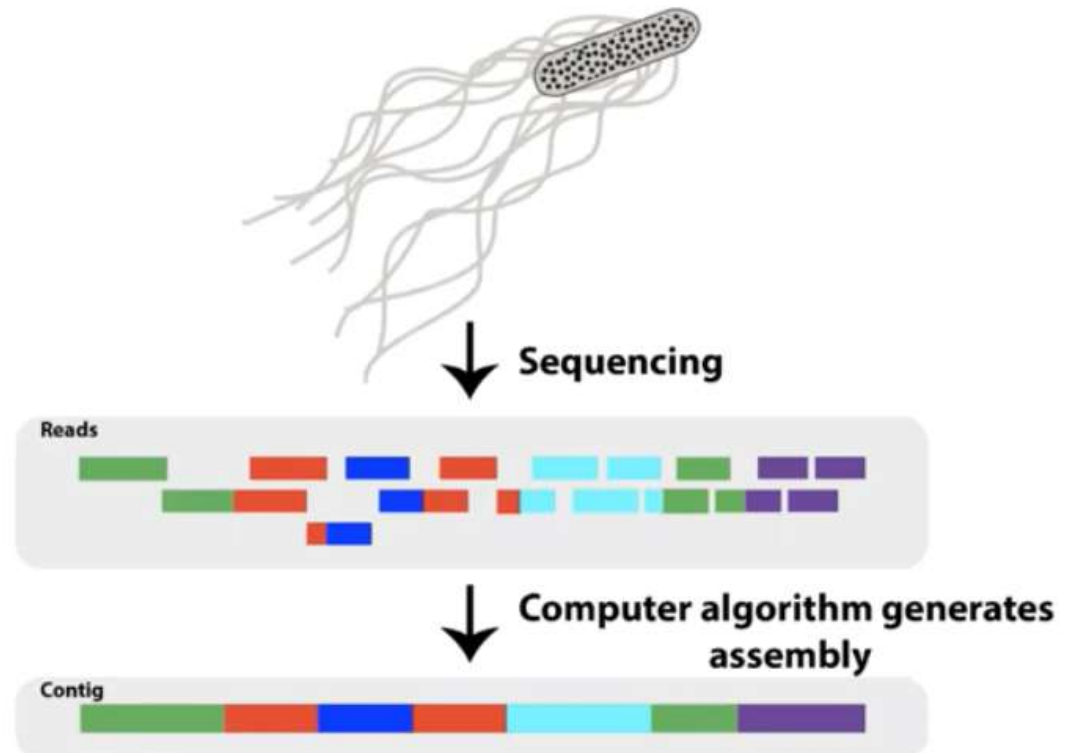


# Types of Assembly

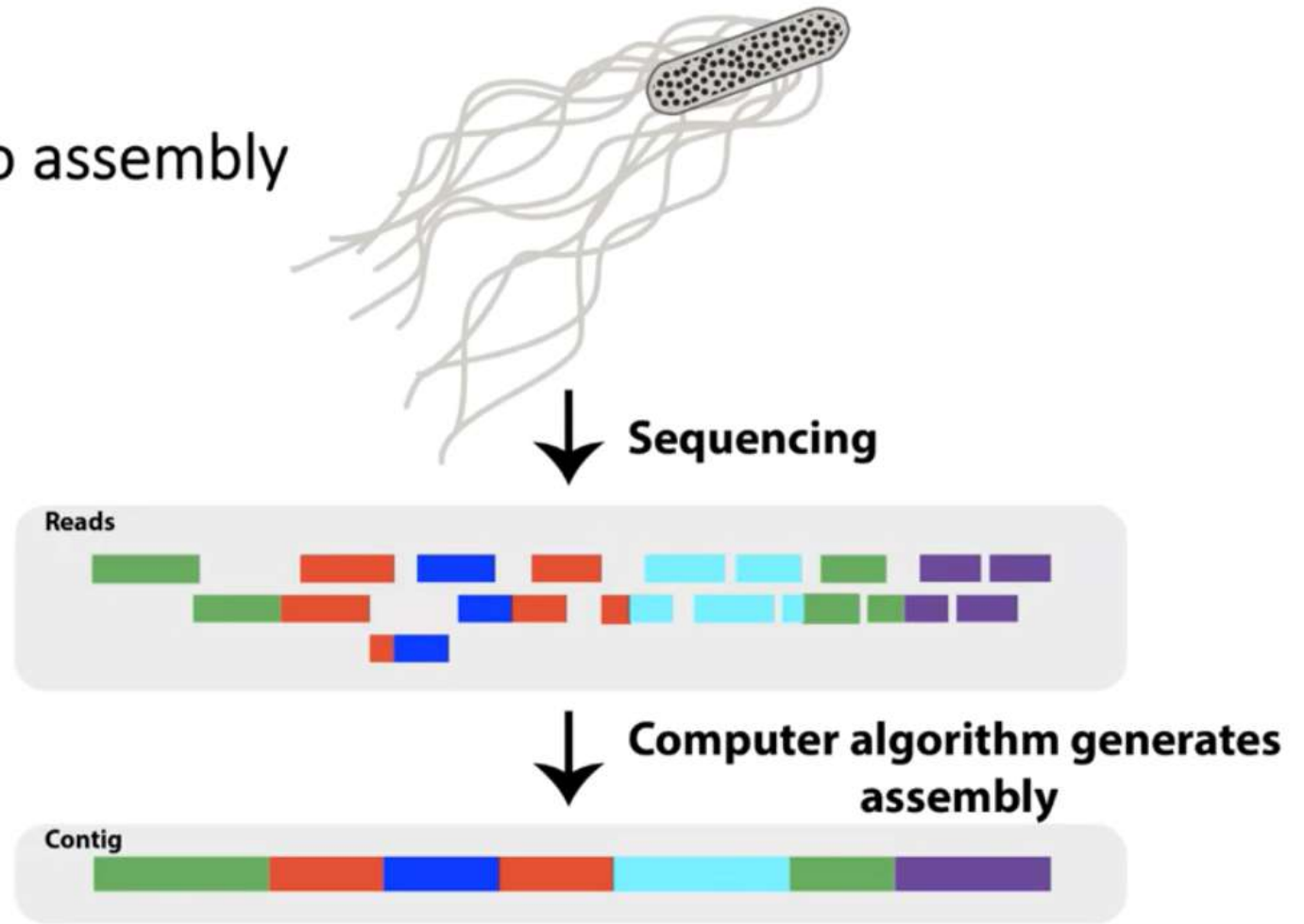
## Reference based



## De novo



PATRIC offers de novo assembly



# Sequencing reads

- Illumina
  - Average read length typically 50-300 nucleotides per read
- Ion Torrent
  - Average read length of approximately 400 nucleotides per read
- PacBio
  - Average read length 10-16 kb nucleotides per read, but sometimes shorter
- Nanopore
  - Read length ranges from 500 to 2.3 Mb nucleotides per read



# Sequencing strategies

- Unicycler
  - Can be used for hybrid assembly
- SPAdes
  - Short reads only
- Canu
  - Long-reads
- metaSPAdes
  - Metagenomes
- plasmidSPAdes
  - Assembling plasmids from whole genome sequencing data
- Multiple Displacement Amplification (MDA)
  - Whole genome amplification prior to single-cell sequencing
  - SPAdes strategy

**Parameters** ⓘ

ASSEMBLY STRATEGY

Auto

**Auto**

Unicycler

SPAdes

Canu

metaSPAdes

plasmidSPAdes

MDA (single-cell)

# Hybrid genome assembly

- Combines long and short reads to assemble a genome
  - Unicycler
  - Can first do Canu, and then polish with short reads using the pylon iterations

# Polishing the assembly

- Tries to fix small errors in the assemblies
  - Goes through the reads to correct base pair errors or short regions
- Racon is for long reads
- Pilon is for short reads

**Parameters** ⓘ

ASSEMBLY STRATEGY

Unicycler

OUTPUT FOLDER

↓<sup>A</sup><sub>Z</sub>

OUTPUT NAME

*Output Name*

ADVANCED ◀

TRIM READS BEFORE ASSEMBLY

False

RACON ITERATIONS PILON ITERATIONS

2 2

MIN. CONTIG LENGTH MIN. CONTIG COVERAGE

300 5

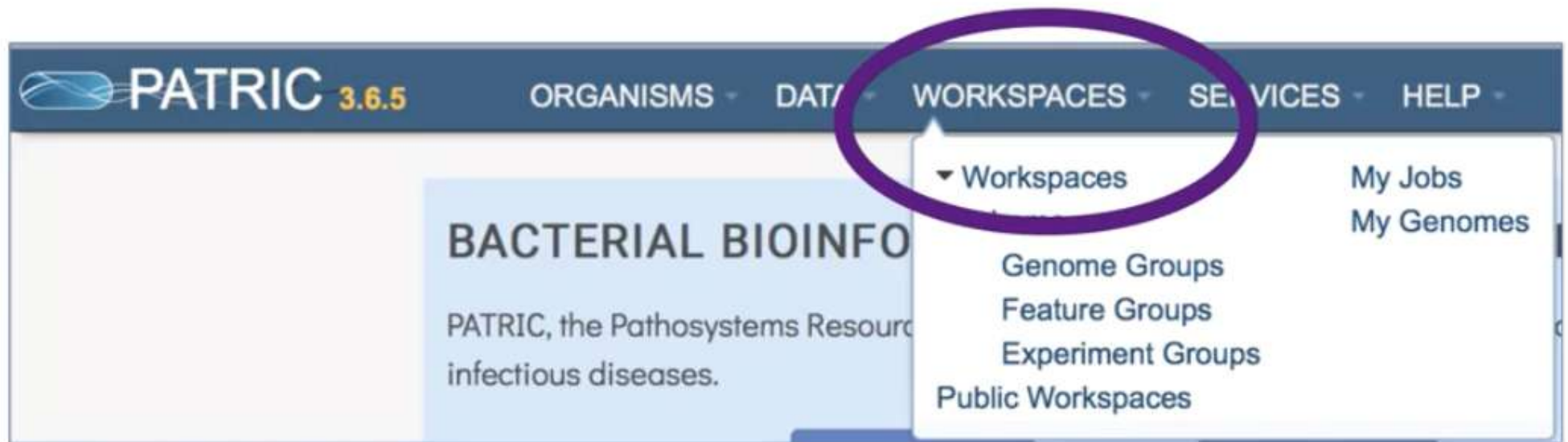


# Assembly course

- Several short instructional videos
- Question follows each video
- Location of test data

# Assembly course

- Several short instructional videos
- Question follows each video
- Location of test data





# Assembly Homework Assignment

Platform	Strategy	Name	Time to assemble	Short read coverage	Total contigs	Largest contig	Total length	#Ns per 100 kbp	Notes
Ion Torrent	Unicycler								
	SPAdes								
	Canu								
	metaSPAdes								
	plasmid SPAdes								
	MDA								
Illumina	Unicycler								
	SPAdes								
	Canu								
	metaSPAdes								
	plasmid SPAdes								
	MDA								