**Project 6**

**Phage Genome Analysis (from raw reads)**

**Objective:**

Assemble your phage genome from raw reads, annotate genes, place the phage in a simple phylogenetic context, and evaluate genes relevant for therapy (e.g., lytic enzymes) and biosafety (no ARGs/virulence).

**Data source**

* Raw paired FASTQ files (provided by instructor) — e.g., sample\_R1.fastq.gz and sample\_R2.fastq.gz
* Reference resources (online): NCBI RefSeq viral database, Phage-specific databases (optional), CARD (for ARGs), VFDB (for virulence factors)

**Steps**

1. Quality control (QC) of raw reads
2. Trim/clean reads
3. De novo assembly
4. Assembly evaluation & circularity check
5. Genome annotation
6. Search for therapy-relevant genes (endolysins, tail fibers) and safety checks (ARGs/virulence)
7. Simple phylogenetic placement using one conserved gene (e.g., terminase large subunit) or whole-genome BLAST hits
8. Short report and presentation

**Tools**

* QC / trimming: FastQC, fastp (fast, all-in-one)
* Assembly: SPAdes (recommended) or Unicycler (if circularization is desired)
* Assembly QC: QUAST, Bandage (visualize assembly graph)
* Annotation: Prokka (command line) and/or **RAST** (web)
* Phage-specific gene callers (optional advanced): PHANOTATE
* Search/BLAST: blastn / blastp against NCBI or local databases
* ARG / virulence screening: ABRicate (with CARD, VFDB) or BLAST against CARD/VFDB
* Visualization & phylogeny: MEGA (GUI), Clustal Omega (online) for MSA, IQ-TREE or MEGA for trees
* Optional visualization: EasyFig or Mauve for genome comparisons; Artemis for viewing features
* Environment: conda recommended to install tools.