Identification of viral contigs using RNA-seq data in tomato (*Solanum lycopersicum* L.)

Introduction: Tomato (*Solanum lycopersicum* L.) is a night shade family crop with high economic value for its produce. However, the tomato production is restricted by numerous diseases and, of all pathogens, viruses are known to significantly affect its production. Earlier virus detection techniques were traditional targeted detection techniques including ELISA, PCR, and Sanger sequencing which require prior information on viral genomes or information for serological properties of viral species. Presently, the lower cost of high-throughput sequencing (HTS) and the availability of the improved data analysis tools has made possible the study of a whole community of viruses (i.e., viromes), including unknown ones. This advancement has escaped challenges of targeted detection of plant viruses and contributed useful ecological and epidemiological insights.

Research question:

How to identify viral contigs using RNA-seq data derived from infected tomato samples?

Objectives:

1. Identification of viral contigs present in the RNA-seq data of infected tomato samples
2. Detection of any novel or uncharacterized viruses if present in the RNA-seq data

Methodology:

1. **Data preprocessing:** The RNA-Seq reads will be subjected to quality control to remove low-quality sequences and adapter contaminants, ensuring that only high-quality data are retained for downstream analysis.
2. **Assembly of reads:** The reads will be assembled into contigs using a *de novo* assembly approach. This step will reconstruct potential viral genomes or fragments from the short RNA-Seq reads.
3. **Viral Contig Identification**: The assembled contigs will be screened for viral sequences by comparing them against a reference database like NCBI.
4. **Functional Annotation and Analysis**: Identified viral contigs will be annotated to determine their functional roles.

Expected outcomes: Identification of known and unknown viruses from the RNA-seq data.