

Project Proposal

Paslay

2024-09-11

Introduction - study species/system and relevant background

In studying plant viruses, there have been numerous studies whereby researchers investigate available HTS data for the presence of known or novel viruses. Characterization of these viruses is critical for future disease prevention and rapid development of diagnostic methods.

According to NCBI, there is a large amount of data on roses (both RNA and DNA). There may be many viruses that are novel that have not yet been characterized given the current study was focusing on something other than virus discovery and characterization.

I personally think it could be very useful to develop a pipeline/script for downloading HTS data and using other tools to create a system that can conduct trimming/QC, de novo assembly, reference mapping, blasting of contigs, and output a results table as the final product.

Research question(s)

1. Are there viruses that are capable of infecting roses that have not yet been reported, but have been sequenced “serendipitously” by HTS methods?
2. Are there novel viruses that have not yet been characterized but were sequenced as a “byproduct” of some other experiment?

Methods - general description of the analysis approaches you will use.

1. Collect SRA data (transcriptomic) for roses
2. Quality check and trimming
3. Mapping raw reads to reference genome (rose genome)
4. Collecting unmapped reads
5. De novo assembly of unmapped reads
6. BLAST analysis to find plant virus contigs

Predictions/expected outcomes

I am hopeful to develop a useful tool for analyzing our own HTS data. In addition, I am hopeful to use this tool to “screen” available HTS data for the presence of plant viruses in roses. In practice, this could be used for any plant host to determine if there are novel or unreported viruses from a given region of the world. Ideally, it would be useful to create a pipeline where we can simply input the raw fastq files and output the results from the BLAST search.