

# A very cool project to start

Ricardo I. Alcalá-Brseño

## Abstract

Viruses plays different roles in plant ecosystems, and viruses are cool, that's why we studied them.

## 1 Introduction

The study of viruses started by characterizing tobacco mosaic virus, genus *Tobamovirus*, family *Virgaviridae* in tobacco plants. Since then, thousands of virus species have been described in plants and other organisms such as fungi and oomycetes, including other phyla, and not to mention other domains [12, 5]. However, very little is known about the diversity of viruses in the United States, and virus epidemics are understudied in the world.

## 2 Background

### 2.1 Viral Isolation

Extraction methods are specific for each group of viruses. These techniques can target naked viruses, such *Amalgavirus* and *Endornavirus*, which are double-stranded RNA viruses [2, 3]. Other methods target viruses surrounded by capsid proteins (encapsidated), either single-stranded RNA genomes, which are most plant viruses. And circular DNA viruses, single-stranded genomes in the families *Geminviridae* and *Nanoviridae* [4], and double-stranded DNA viruses in the *Caulimoviridae* family –or pararetroviruses– some genera in this family can integrate into the host genome [14].

The most common method for virus sequencing is Illumina, constructing RNA or DNA libraries for single or paired-end reads (I prefer PE X 150 or 250 nt) [2, 3, 1]. I can implement standard methods for determining read quality, and tag removal can be used. However, some sequencing results yield double-stranded nucleic acids that are expected due to the nature of the purification methods. Part of the sequence filtering is implemented in the bioinformatic pipeline.

## 2.2 Viral Metagenomics

The implementation of viral metagenomics allows the characterization of viruses using high-throughput sequencing approaches and bioinformatic tools [11, 13]. Several approaches to sequence viruses have been implemented by bulking samples of different species and sequence them (plant metagenomics) to several species from ecological and agricultural importance sampled in bulks (agroeco-genomics) and individual hosts (eco-genomics) [13]. Several viral assembly pipelines have been developed and tested for virus identification [10]. Most programs perform well for abundant viruses, and a couple of software are more sensitive to less abundant species. Several custom pipelines can be implemented according to the virome characteristic and research needs. For example, in the papaya orchard virome [1], a full-length genome was the minimum threshold to count for a virus species. Other approaches consist in assigning viral operational taxonomic units (vOTUs) to assign viral categories [8], or others that implement semi-supervised machine learning that assigns species according to the International Committee for the Taxonomy of Viruses (ICTV) demarcation criteria of virus species (Alcalá-Briseño et al., manuscript in prep).

## 2.3 Virus Characterization and Taxonomy

The plant samples processed with HTS and bioinformatic pipelines help to identify viruses present in the host. The virus identified can be represented as a super contig representing a full-length virus genome or multiple contigs that can be scaffolded to generate a full-length virus genome. However, multiple contigs often corresponding to a virus species can not generate a full length and are represented as partial sequences that are usually problematic to classify (Fig 1).

Fortunately, characterizing full-length genome sequences of viruses is relatively straightforward. Most methods utilize BLAST searches to identify sequences homology to viruses, narrowing it down to one of the 65 plant virus genera described till now [7, 9]. Using sequence data, the ICTV determines the virus species demarcation criteria based on pairwise similarities and/or phylogenetics using sequence data [9]. However, other physicochemical and biological characteristics are considered [7, 9]. Genome and gene phylogenetic relationship dictates virus taxonomical classification for species, genus, family, and order. Until recently, that virus classification was updated to seven realms [6].

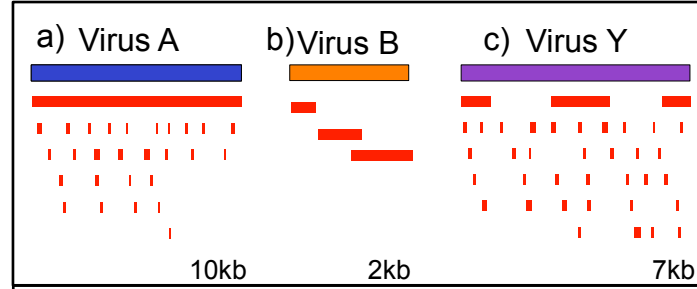


Figure 1: An example of contigs partition. a) Super contig assembled, b) several contigs assembled allowing to generate a scaffold, c) several contigs that does not allow to generate a scaffold

## References

- [1] ALCALÁ-BRISEÑO, R. I., CASARRUBIAS-CASTILLO, K., LÓPEZ-LEY, D., GARRETT, K. A., AND SILVA-ROSALES, L. Network analysis of the papaya orchard virome from two agroecological regions of chiapas, mexico. *Msystems* 5, 1 (2020), e00423–19.
- [2] ALCALÁ-BRISEÑO, R. I., COŞKAN, S., LONDOÑO, M. A., AND POLSTON, J. E. Genome sequence of southern tomato virus in asymptomatic tomato ‘sweet hearts’. *Genome Announc.* 5, 7 (2017), e01374–16.
- [3] ALCALÁ-BRISEÑO, R. I., OKADA, R., HERRERA, F., AND VALVERDE, R. A. A novel endornavirus isolated from cluster bean (cyamopsis tetragonoloba). *Archives of virology* 163, 8 (2018), 2279–2282.
- [4] BOUKARI, W., ALCALÁ-BRISEÑO, R. I., KRABERGER, S., FERNANDEZ, E., FILLOUX, D., DAUGROIS, J.-H., COMSTOCK, J. C., LETT, J.-M., MARTIN, D. P., VARSANI, A., ET AL. Occurrence of a novel mastrevirus in sugarcane germplasm collections in florida, guadeloupe and réunion. *Virology journal* 14, 1 (2017), 1–8.
- [5] DOLJA, V. V., KRUPOVIC, M., AND KOONIN, E. V. Deep roots and splendid boughs of the global plant virome. *Annual Review of Phytopathology* 58, 1 (2020), 23–53.
- [6] GORBALENYA, A. E., KRUPOVIC, M., MUSHEGIAN, A., KROPINSKI, A. M., SIDDELL, S. G., VARSANI, A., ADAMS, M. J., DAVISON, A. J., DUTILH, B. E., HARRACH, B., HARRISON, R. L., JUNGLEN,

- S., KING, A. M. Q., KNOWLES, N. J., LEFKOWITZ, E. J., NIBERT, M. L., RUBINO, L., SABANADZOVIC, S., SANFAÇON, H., SIMMONDS, P., WALKER, P. J., ZERBINI, F. M., KUHN, J. H., AND INTERNATIONAL COMMITTEE ON TAXONOMY OF VIRUSES EXECUTIVE COMMITTEE. The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. *Nature Microbiology* 5, 5 (2020), 668–674.
- [7] KING, A., LEFKOWITZ, E., ADAMS, M., AND CARSTENS, E. *Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses*. Academic Press.
- [8] LEFEBVRE, M., THEIL, S., MA, Y., AND CANDRESSE, T. The viranot pipeline: A resource for automated viral diversity estimation and operational taxonomy units assignment for virome sequencing data. *Phytobiomes Journal* 3, 4 (2019), 256–259.
- [9] LEFKOWITZ, E. J., DEMPSEY, D. M., HENDRICKSON, R. C., ORTON, R. J., SIDDELL, S. G., AND SMITH, D. B. Virus taxonomy: the database of the international committee on taxonomy of viruses (ICTV). *Nucleic Acids Research* 46 (2018), D708–D717.
- [10] MASSART, S., CHIUMENTI, M., DE JONGHE, K., GLOVER, R., HAEGEMAN, A., KOLONIUK, I., KOMÍNEK, P., KREUZE, J., KUTNJAK, D., LOTOS, L., MACLOT, F., MALIOGKA, V., MAREE, H. J., OLIVIER, T., OLMOS, A., POOGGIN, M. M., REYNARD, J.-S., RUIZ-GARCÍA, A. B., SAFAROVA, D., SCHNEEBERGER, P. H. H., SELA, N., TURCO, S., VAINIO, E. J., VARALLYAY, E., VERDIN, E., WESTENBERG, M., BROSTAU, Y., AND CANDRESSE, T. Virus detection by high-throughput sequencing of small rnas: Large-scale performance testing of sequence analysis strategies. *Phytopathology*® 109, 3 (2019), 488–497. PMID: 30070618.
- [11] ROOSSINCK, M. J. Metagenomics of plant and fungal viruses reveals an abundance of persistent lifestyles. *Frontiers in Microbiology* 5 (2015), 767.
- [12] ROOSSINCK, M. J. Viruses in the phytobiome. *Current Opinion in Virology* 37 (2019), 72–76.
- [13] ROOSSINCK, M. J., MARTIN, D. P., AND ROUMAGNAC, P. Plant virus metagenomics: Advances in virus discovery. *Phytopathology* 105, 6 (2015), 716–727.

- [14] SAAD, N., ALCALÁ-BRISEÑO, R., POLSTON, J., OLMSTEAD, J., VARSANI, A., AND HARMON, P. Blueberry red ringspot virus genomes from florida inferred through analysis of blueberry root transcriptomes. *Scientific reports* 10, 1 (2020), 1–10.