

# An automated strategy for viral diversity estimation and OTUs assignation implemented in the VirAnnot pipeline

Marie Lefebvre, Yuxin Ma, Sébastien Theil, Thierry T. Candresse

▶ **To cite this version:**

Marie Lefebvre, Yuxin Ma, Sébastien Theil, Thierry T. Candresse. An automated strategy for viral diversity estimation and OTUs assignation implemented in the VirAnnot pipeline. 17. Rencontres de Virologie Végétale (RVV 2019), Jan 2019, Aussois, France. 2019. hal-02737922

**HAL Id: hal-02737922**
**https://hal.inrae.fr/hal-02737922**

Submitted on 2 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

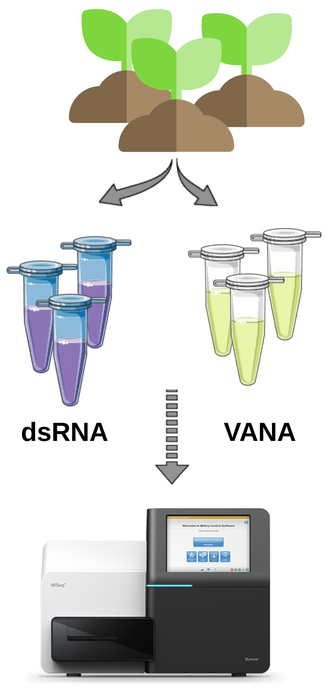
# An automated strategy for viral diversity estimation and OTUs assignation implemented in the VirAnnot pipeline

Marie Lefebvre, Yuxin Ma, Sébastien Theil, Thierry Candresse
UMR 1332 Biologie du Fruit et Pathologie, INRA, Univ. Bordeaux, CS20032, 33882 Villenave d'Ornon cedex, France
marie.lefebvre@inra.fr

## 1. Introduction

**Metagenomics** can access the genetic content and diversity of entire microbial communities relying on high-throughput sequencing (HTS) and on bioinformatic analyses. Whereas metagenomics face identification, storage and computational challenges, viral metagenomic is confronted to specific taxonomic assignation and diversity evaluation difficulties. Unlike fungi or bacteria for which universally shared genes (ITS, 16S ribosomal RNA) can be used to define **Operational Taxonomic Units** (OTUs) through a clustering approach, no such universally shared pattern exists for viruses. Currently there is no universally accepted strategy or tool to define OTUs and evaluate viral species richness from virome data.
We have developed an automated routine adressing these issues and integrated it in our virome annotation bioinformatic pipeline, **virAnnot**.

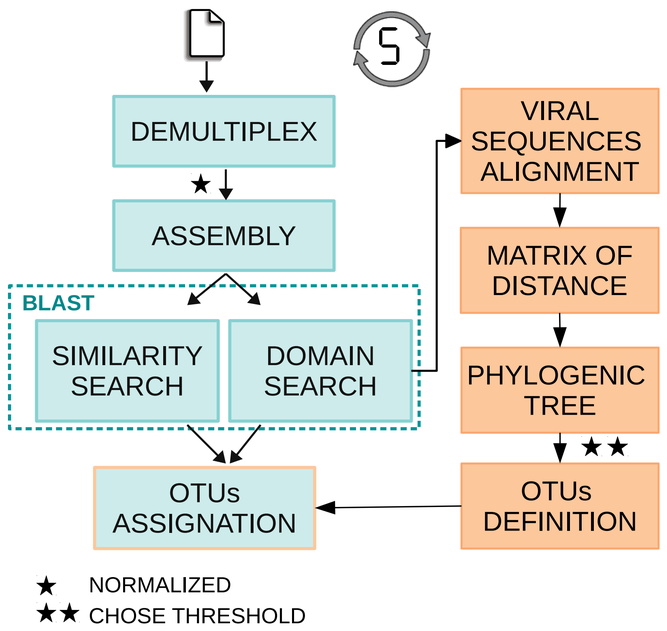
## 2. Material




Talk n°02 by Yuxin Ma.

## 3. Methods

The VirAnnot pipeline integrates standard bioinformatic tools.




- The Domain Search (RPS-Blast) retrieves contigs belonging to viral conserved domain *e.g.* RdRp domain.
- Contigs identity is validated by BlastX results. Motives containing regions of validated contigs are then aligned.
- Then for each alignment corresponding to a protein domain, a distance matrix and a taxonomic tree are generated.
- A clustering approach based on a distance threshold is finally used to define OTUs.

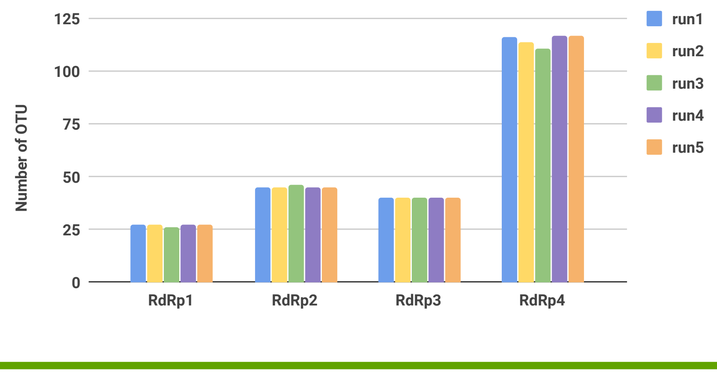
➡ To verify the **repeatability** of the analysis we ran it 5 times independently on two datasets.

➡ We routinely use a **10% distace threshold**. To validate it we compared the assignment of OTU and the ICTV taxonomy [1] at the inter-species level for several viral families.

## 4. Results

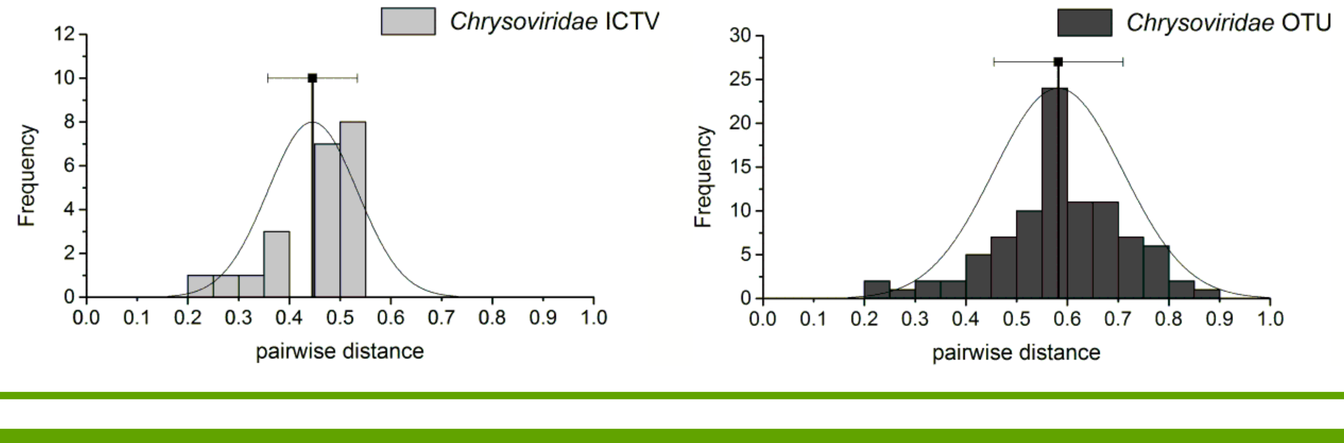
### Repeatability

- Reads assembly and OTUs definition were repeated 5 times.
- The number of OTUs for each of the four RdRp domain is shown.
- The results show a good stability of the results. The limited variations are likely due to the known variability of the assembly and clustering processes.




### Validity

- Here we compare amino-acid (aa) pairwise distances in the short conserved region around the RdRp motives of virAnnot OTUs and of formal species listed in ICTV.
- In the *Chrysoviridae* family, the minimum aa distance was found to be 22% for the ICTV species, whereas it was 23% for the OTUs.
- This comparison was performed for 6 other families with similar good fit between OTUs and ICTV species, supporting the meaningfullness of the 10% threshold but this parameter can be adjusted at will in virAnnot.




## Availability & References

Pipeline available at :
**https://github.com/marieBvr/virAnnot**

[1] E. J. Lefkowitz et al. "Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV)". *Nucleic Acids Research* 46.D1 (2018), pp. D708–D717.

## 5. Conclusions

1. VirAnnot allows the automated evaluation of viral species richness in Metavirome data and therefore an easy and direct comparison between samples.
2. We confirmed the stability and repeatability of the virAnnot species richness and OTUs assignation.
3. The 10% divergence threshold for OTUs definition provides results that seem to mimic in many families the ICTV taxonomy, allowing to use OTUs as a proxy to validate species.
4. Use of higher or lower threshold values may allow the definition of OTUs representing different taxonomic levels (inter or intra-specific).