

# Búsqueda de Tobamovirus en la base de datos Serratus

<https://serratus.io/explorer/rdrp> (<https://serratus.io/explorer/rdrp>)

Seleccionar Family, escribir/seleccionar 'Virgaviridae'

Seleccionar view matches

Resultan en más de 900!

## Restringir la búsqueda a las especies del género Tobamovirus

Las especies dentro del género son (<https://en.wikipedia.org/wiki/Tobamovirus>  
(<https://en.wikipedia.org/wiki/Tobamovirus>))

Bell pepper mottle virus (BPeMV)  
Brugmansia mild mottle virus  
Cactus mild mottle virus (CMMoV)  
Clitoria yellow mottle virus  
Cucumber fruit mottle mosaic virus  
Cucumber green mottle mosaic virus (CGMMV)  
Cucumber mottle virus  
Frangipani mosaic virus (FrMV)  
Hibiscus latent Fort Pierce virus (HLFPV)  
Hibiscus latent Singapore virus (HLSV)  
Kyuri green mottle mosaic virus  
Maracuja mosaic virus (MarMV)  
Obuda pepper virus (ObPV)  
Odontoglossum ringspot virus (ORSV)  
Opuntia chlorotic ringspot virus  
Paprika mild mottle virus  
Passion fruit mosaic virus  
Pepper mild mottle virus (PMMoV)  
Plumeria mosaic virus  
Rattail cactus necrosis-associated virus (RCNaV)  
Rehmannia mosaic virus  
Ribgrass mosaic virus (HRV)  
Sammons s Opuntia virus (SOV)  
Streptocarpus flower break virus  
Sunn-hemp mosaic virus (SHMV)  
Tobacco latent virus  
Tobacco mild green mosaic virus  
Tobacco mosaic virus (TMV)  
Tomato brown rugose fruit virus (ToBRFV)  
Tomato mosaic virus (ToMV)  
Tomato mottle mosaic virus  
Tropical soda apple mosaic virus  
Turnip vein-clearing virus (TVCV)  
Ullucus mild mottle virus  
Wasabi mottle virus (WMoV)  
Yellow tailflower mild mottle virus  
Youcai mosaic virus (YoMV) aka oilseed rape mosaic virus (ORMV)  
Zucchini green mottle mosaic virus

## Contigs de virus que tienen RdRP

Descargar de <https://github.com/ababaian/serratus/wiki/Viral-contigs-containing-RdRP>  
(<https://github.com/ababaian/serratus/wiki/Viral-contigs-containing-RdRP>)

Específicamente de:

[https://serratus-public.s3.amazonaws.com/rdrp\\_contigs/rdrp\\_contigs.tar.gz](https://serratus-public.s3.amazonaws.com/rdrp_contigs/rdrp_contigs.tar.gz) ([https://serratus-public.s3.amazonaws.com/rdrp\\_contigs/rdrp\\_contigs.tar.gz](https://serratus-public.s3.amazonaws.com/rdrp_contigs/rdrp_contigs.tar.gz))

Verificar la presencia virus en los contigs que contienen RdRP en Serratus

```
grep 'Tobamovirus' '/media/eugenia/megalodon1/Virus/rdrp_contigs/rdrp_contigs.tsv' | wc -l  
#12626
```

Visualizar lista con less

```
grep 'Tobamovirus' '/media/eugenia/megalodon1/Virus/rdrp_contigs/rdrp_contigs.tsv' | less -S
```

Obtener solo lista de 'accession numbers' con awk (imprime segunda columna) y sort -u (ordena e imprime solo registros únicos)

```
grep 'Tobamovirus' '/media/eugenia/megalodon1/Virus/rdrp_contigs/rdrp_contigs.tsv' | awk -F '\t' '{  
print $2}' | sort -u > acc_num_uni.txt
```

Cuántas líneas hay = cuantos registros únicos hay

```
grep 'Tobamovirus' '/media/eugenia/megalodon1/Virus/rdrp_contigs/rdrp_contigs.tsv' | awk -F '\t' '{  
print $2}' | sort -u | wc -l  
#5869
```

Verificar cuántas especies están representadas

```
grep 'Tobamovirus' '/media/eugenia/megalodon1/Virus/rdrp_contigs/rdrp_contigs.tsv' | awk -F '\t' '{  
print $19}' | sort -u | wc -l  
#20
```

Cuáles son las especies representadas

```
grep 'Tobamovirus' '/media/eugenia/megalodon1/Virus/rdrp_contigs/rdrp_contigs.tsv' | awk -F '\t' '{  
print $19}' | sort -u  
#.  
#Cucumber green mottle mosaic virus  
#Hibiscus latent Fort Pierce virus  
#Hibiscus latent Singapore virus  
#Lentinula edodes tobamo-like virus 1  
#Odontoglossum ringspot virus  
#Pepper mild mottle virus  
#Rattail cactus necrosis-associated virus  
#Ribgrass mosaic virus  
#Tobacco mild green mosaic virus  
#Tobacco mosaic virus  
#Tomato brown rugose fruit virus  
#Tomato mosaic virus  
#Tomato mottle mosaic virus  
#Tropical soda apple mosaic virus  
#Turnip vein-clearing virus  
#Wasabi mottle virus  
#Watermelon green mottle mosaic virus  
#Youcai mosaic virus  
#Zucchini green mottle mosaic virus
```

Crear archivo con rdrp contigs de Tobamovirus en Serratus

```
head -1 '/media/eugenia/megalodon1/Virus/rdrp_contigs/rdrp_contigs.tsv' > rdrp_contigs_tobamovirus  
.txt  
grep 'Tobamovirus' '/media/eugenia/megalodon1/Virus/rdrp_contigs/rdrp_contigs.tsv' >> rdrp_contigs  
_tobamovirus.txt
```

# Obtener metadatos

<https://www.ncbi.nlm.nih.gov/sra/docs/srdownload/#download-metadata-associated-with>  
(<https://www.ncbi.nlm.nih.gov/sra/docs/srdownload/#download-metadata-associated-with>)

Utilizar SRA tools y entrez direct. Crear ambiente en conda para instalarlos.

```
conda create --name serratus
conda activate serratus
conda install -c bioconda sra-tools
#instalar sra tools
conda install -c bioconda sra-tools
#instalar entrez
conda install -c bioconda entrez-direct
```

Utilizar esearch como en este ejemplo:

```
esearch -db sra -query SRR10931999 | efetch -format runinfo
```

Crear una lista con los accession numbers de los archivos que tienen Tobamovirus. Para ello el archivo rdrp\_contigs\_tobamovirus.txt se importó en excel y se copió la segunda columna - la que contiene los accession numbers (get\_metadata\_SRA\_batch\_construction.xls). Esta se utilizó para crear un script para obtener los metadatos (get\_metadata\_SRA\_batch.sh).

correr el script para obtener metadatos de todas las secuencias donde se ha detectado Tobamovirus

```
bash get_metadata_SRA_batch.sh
#resultados en acc_uni_metadata.txt
```

Crear y correr un script para solo quedarnos con metadata sin repeticiones

```
bash get_metadata_SRA_unique_runs_batch.sh
#requiere acc_uni_metadata.txt. Resultados en: acc_unique_runs_metadata.txt
```

Este archivo se puede importar a excel para:

1. Conocer dónde se detectó Tobamovirus, de acuerdo a la columna 'ScientificName'
2. Cuántos y qué tipos distintos de muestras se analizaron, también de acuerdo a la columna 'ScientificName' e.g Solanum lycopersicum, wetland metagenome, wastewater metagenome, salt marsh metagenome, riverine metagenome, lake water metagenome, freshwater metagenome, aquatic metagenome, etc.
3. Hacer un archivo de metadatos solo con accession numbers de ambientes de interés agua residual y aguas superficiales (ríos, lagos). Este servirá para poder extraer los contigs de Tobamovirus que se encuentran en estos ambientes.
4. Tratar de identificar origen geográfico