

Some links: Serratus [Website](https://serratus.io/toolkit), [Article](https://www.nature.com/articles/s41586-021-04332-2), [Documentation](https://github.com/ababaian/serratus/wiki/SQL-Schema)

The website can be used to explore starting from Rdrp Palm-id of a known or newly discovered virus. But exploring the database allows us to explore in other directions:

1. Explore the host range of a specific virus
2. Find (non-documented) known virus on a specific plant species/type
3. Try to see if a new virus was previously present in SRA (and detected by serratus).

Accessing the DB: you can use the PGadmin interface <https://www.pgadmin.org/>

Several tables compose the database. Two are very important for our search, the rsequence (corresponding to the lowest level of sequence matches for RdRp search) and srarun (description of SRA dataset).

By joining (combining) the two tables, we can obtain information on the SRA that has RdRp hits filtered on the plant or virus.

A selection of the interesting information to keep is done (fields name after SELECT). Then the two tables are joined (after FROM) using a right join strategy (more [information](https://commons.wikimedia.org/wiki/File:SQL_Joins.svg) on alternative join) based on run/run\_id (SRA unique ID). Finally, a filter (after WHERE) is applied to the SRA dataset scientific name and the number of reads (an arbitrary limit of 100 reads to avoid having too many false positive results). You may note the % after tomato; it means that the scientific name should start with tomato(case sensitive) and can have any characters after (to allow all tomato viruses).

1. Explore the host range of a specific virus

tomato virus

SELECT A.run\_id, A.family\_name, A.virus\_name, B.bio\_project, B.sample, B.bio\_sample, B.library\_name, B.library\_strategy, B.library\_source, B.scientific\_name, A.score, A.percent\_identity, A.depth, A.n\_reads

FROM rsequence A RIGHT JOIN srarun B ON run\_id=run

WHERE A.virus\_name like 'tomato%' and A.n\_reads>=100

SELECT A.run\_id, A.family\_name, A.virus\_name, B.bio\_project, B.sample, B.bio\_sample, B.library\_name, B.library\_strategy, B.library\_source, B.scientific\_name, A.score, A.percent\_identity, A.depth, A.n\_reads

FROM rsequence A RIGHT JOIN srarun B ON run\_id=run

WHERE A.virus\_name like 'tomato\_brown\_rugose\_fruit\_virus' and A.n\_reads>=100

SELECT A.run\_id, A.family\_name, A.virus\_name, B.bio\_project, B.sample, B.bio\_sample, B.library\_name, B.library\_strategy, B.library\_source, B.scientific\_name, A.score, A.percent\_identity, A.depth, A.n\_reads

FROM rsequence A RIGHT JOIN srarun B ON run\_id=run

WHERE A.virus\_name like 'little\_cherry%' and A.n\_reads>=100

search based on family

SELECT A.run\_id, A.family\_name, A.virus\_name, B.bio\_project, B.sample, B.bio\_sample, B.library\_name, B.library\_strategy, B.library\_source, B.scientific\_name, A.score, A.percent\_identity, A.depth, A.n\_reads

FROM rsequence A RIGHT JOIN srarun B ON run\_id=run

WHERE A.family\_name like 'Closteroviridae' and A.n\_reads>=100

Ian Addams (maize lethal disease) the two viruses are needed to have the disease (choice of or/and in WHERE is important!):

SELECT A.run\_id, A.family\_name, A.virus\_name, B.bio\_project, B.sample, B.bio\_sample, B.library\_name, B.library\_strategy, B.library\_source, B.scientific\_name, A.score, A.percent\_identity, A.depth, A.n\_reads

FROM rsequence A RIGHT JOIN srarun B ON run\_id=run

WHERE (A.virus\_name like 'maize\_chlorotic\_mottle\_virus' or A.virus\_name like 'sugarcane\_mosaic\_virus') and A.n\_reads>=100

1. Find (non-documented) known virus on a specific plant species/type

Virus in apple

SELECT A.run\_id, A.family\_name, A.virus\_name, B.bio\_project, B.sample, B.bio\_sample, B.library\_name, B.library\_strategy, B.library\_source, B.scientific\_name, A.score, A.percent\_identity, A.depth, A.n\_reads

FROM rsequence A RIGHT JOIN srarun B ON run\_id=run

WHERE B.scientific\_name like 'Malus%' and A.n\_reads>=100

SELECT A.run\_id, A.family\_name, A.virus\_name, B.bio\_project, B.sample, B.bio\_sample, B.library\_name, B.library\_strategy, B.library\_source, B.scientific\_name, A.score, A.percent\_identity, A.depth, A.n\_reads

FROM rsequence A RIGHT JOIN srarun B ON run\_id=run

WHERE B.scientific\_name like 'Prunus avium' and A.n\_reads>=100

1. Try to see if a new virus was previously present in SRA (and detected by serratus).

! the current request is not giving the expected result => problem on consecutive join !!

we probably need a subquery system: https://www.w3resource.com/sql/subqueries/understanding-sql-subqueries.php

SELECT A.run\_id, A.family\_name, A.virus\_name, B.bio\_project, B.sample, B.bio\_sample, B.library\_name, B.library\_strategy, B.library\_source, B.scientific\_name, C.palm\_id, A.score, A.percent\_identity, A.depth, A.n\_reads

FROM rsequence A JOIN srarun B ON A.run\_id=B.run JOIN palm\_sra C on A.run\_id=C.run\_id

WHERE C.palm\_id like 'u245' and A.n\_reads>=100