Stop worrying and use R to query EPPO

Michał Czyż 2021-06-12

Search

Copy-paste

Go to the next link

Spreadsheet

Do not show this to your boss!

Operating system	Per-minute rate
Linux	$0.008
macOS	$0.08
Windows	$0.016

https://github.com

Use scripting language to do boring stuff for you.

Enjoy your time and coffe!

As it is appealing it takes time and skills to write code.

Not the most effective thing to do, for rare or *one-time* tasks.

The more often code is used the more expensive it is to maintain. Thus, most effort should be put into development phase.

Don't worry! I will take burden from your shoulders with

Ⓡ Package to make batch queries for EPPO Data Services.

Only most basic skills needed.

Output can be easily processed with Ⓡ or dumped into spreadsheet.

pestr structure

SQLite database tools

- eppo_database_check
- eppo_database_connect
- eppo_database_download
- eppo_tabletools_names
- eppo_names_tables

Authorization token tools

- ×check_eppo_token
- ×create_eppo_token

EPPO Global Database/Data Services tools

- eppo_tabletools_host
- eppo_tabletools_pests
- eppo_tabletools_taxo
- eppo_tabletools_cat

You need:

O SQL

O REST API

O Web Scraping

Initial setup

```r
library("pestr")
eppo_database_download()
eppoDB <- eppo_database_connect(filename = "eppocodes.sqlite")
eppo_token <- create_eppo_token("your_eppo_token")
```

Basic functions

```r
pests_query <- c("Xylella", "suzuki")
hosts_query <- c("Triricum durum", "Abies al")
pests_names <- eppo_names_tables(pests_query)
hosts_names <- eppo_names_tables(hosts_query)
hosts       <- eppo_tabletools_hosts(pests_names, token = eppo_token)
pests       <- eppo_tabletools_pests(hosts_names, token = eppo_token)
category    <- eppo_tabletools_cat(pests_names, token = eppo_token)
taxonomy    <- eppo_tabletools_taxo(pests_names, token = eppo_token)
```

`eppo_names_tables` function return 4 tables: `exists_in_DB`, `not_in_DB`, `pref_names` and `all_associated_names`.

`eppo_tabletools_*` functions family returns always two tables: `long_table` and `compact_table`. The structure of `long_table` is different for different outputs since we are dealing with different data. `compact_table` contains the same data as `long_table` but condensed to one row per `eppocode`.

10

Output structure

```
pests_names$exist_in_DB

##     codeid                          fullname
## 1     9518                 Drosophila suzukii
## 2    10286                  suzuki leafhopper
## 3    10286                      Zygina suzuki
## 4    10286                Erythroneura suzuki
## 5     3243                 Xylella fastidiosa
## 6    60986                            Xylella
## 7     3243 Xylella fastidiosa subsp. fastidiosa
## 8     3243    Xylella fastidiosa subsp. piercei
## 9    70536  Xylella fastidiosa subsp. multiplex
## 10   70537      Xylella fastidiosa subsp. pauca
## 11   70538      Xylella fastidiosa subsp. sandyi
## 12   92965 Xylella fastidiosa subsp. fastidiosa
## 13  101366                  Xylella taiwanensis
## 14  107088      Xylella fastidiosa subsp. tashke
## 15  107089       Xylella fastidiosa subsp. morus
```

```
pests_names$pref_names
```

```
##      codeid                              fullname eppocode
## 1      9518                     Drosophila suzukii   DROSSU
## 2     10286                         Zygina suzuki    ERYTSZ
## 3      3243                     Xylella fastidiosa   XYLEFA
## 4     60986                               Xylella   1XYLEG
## 5     70536   Xylella fastidiosa subsp. multiplex   XYLEFM
## 6     70537       Xylella fastidiosa subsp. pauca   XYLEFP
## 7     70538      Xylella fastidiosa subsp. sandyi   XYLEFS
## 8     92965 Xylella fastidiosa subsp. fastidiosa   XYLEFF
## 9    101366                     Xylella taiwanensis   XYLETA
## 10   107088      Xylella fastidiosa subsp. tashke   XYLEFT
## 11   107089       Xylella fastidiosa subsp. morus   XYLEFO
```

```
head(pests_names$all_associated_names, 10)
```

```
##    codeid                 fullname preferred codelang eppocode
## 1    9518         cherry drosophila         0       en   DROSSU
## 2    9518    drosophile du cerisier         0       fr   DROSSU
## 3    9518        Drosophila suzukii         1       la   DROSSU
## 4   10286         suzuki leafhopper         0       en   ERYTSZ
## 5   10286             Zygina suzuki         1       la   ERYTSZ
## 6   10286       Erythroneura suzuki         0       la   ERYTSZ
## 7    3243          leaf scorch of elm         0       en   XYLEFA
## 8    3243  leaf scorch of mulberry         0       en   XYLEFA
## 9    3243        Xylella fastidiosa         1       la   XYLEFA
## 10  60986                   Xylella         1       la   1XYLEG
```

```
head(hosts$long_table, 10) %>%
    flextable::flextable()  %>%
    flextable::bg(bg = "white", part = "all") %>%
    flextable::width(5:6, 2.5)
```

eppocode	codeid	host_eppocode	idclass	labelclass	full_name
DROSSU	32,628	FRAAN	1	Major host	Fragaria x ananassa
DROSSU	41,521	PRNAR	1	Major host	Prunus armeniaca
DROSSU	41,523	PRNAV	1	Major host	Prunus avium
DROSSU	41,563	PRNDO	1	Major host	Prunus domestica
DROSSU	41,632	PRNPS	1	Major host	Prunus persica
DROSSU	43,667	RUBFR	1	Major host	Rubus fruticosus
DROSSU	65,883	RUBAE	1	Major host	Rubus hedycarpus subsp. armeniacus
DROSSU	43,677	RUBID	1	Major host	Rubus idaeus
DROSSU	43,684	RUBLA	1	Major host	Rubus laciniatus
DROSSU	43,743	RUBSS	1	Major host	Rubus sp.

```
pests$compact_table %>%
    flextable::flextable() %>%
    flextable::bg(bg = "white", part = "all") %>%
    flextable::width(2, 9)
```

eppocode	pests
ABIAL	Experimental: Fusarium circinatum, Melampsora medusae (as Abies), Melampsora medusae f. sp. deltoidis (as Abies); Host: Acleris gloverana (as Abies), Acleris variana (as Abies), Arceuthobium abietinum (as Abies), Arceuthobium douglasii (as Abies), Arceuthobium laricis (as Abies), Bursaphelenchus xylophilus (as Abies), Chionaspis pinifoliae, Chionaspis pinifoliae (as Abies), Choristoneura freemani (as Abies), Choristoneura fumiferana (as Abies), Chrysomyxa abietis (as Abies), Coniferiporia weirii (as Pinaceae), Crisicoccus pini (as Abies), Dendroctonus micans, Dendrolimus sibiricus (as Abies), Dendrolimus spectabilis (as Abies), Dendrolimus superans (as Abies), Dothistroma septosporum, Dryocoetes confusus (as Abies), Gnathotrichus sulcatus (as Pinaceae), Gremmeniella abietina (as Abies), Heterobasidion irregulare (as Abies), Ips amitinus, Ips amitinus (as Abies), Ips sexdentatus, Ips subelongatus (as Abies), Ips typographus, Leptoglossus occidentalis (as Abies), Malacosoma disstria (as Abies), Monochamus alternatus (as Abies), Monochamus marmorator (as Abies), Monochamus obtusus (as Abies), Monochamus saltuarius (as Abies), Monochamus scutellatus (as Abies), Monochamus sutor (as Abies), Monochamus titillator (as Abies), Monochamus urussovi (as Abies), Phacidium coniferarum (as Abies), Phytophthora cinnamomi (as Pinaceae), Phytophthora ramorum, Pissodes castaneus, Polygraphus proximus (as Abies), Sirex ermak (as Abies), Sirex noctilio (as Abies), Tetropium gracilicorne (as Abies), Trichoferus campestris (as Abies); Major host: Chrysomyxa abietis, Monochamus sutor, Neonectria neomacrospora
TRZDU	Host: Atherigona oryzae (as Triticum), Cicadulina mbila (as Triticum), Claviceps purpurea (as Triticum), Delia arambourgi (as Triticum), Ditylenchus dipsaci (as Triticum), Epilachna similis similis (as Triticum), European wheat striate mosaic virus (as Triticum), Listronotus bonariensis (as Triticum), Maize streak virus (as Triticum), Maize stripe tenuivirus (as Triticum), Meloidogyne chitwoodi (as Triticum), Pectobacterium rhapontici (as Triticum), Puccinia graminis (as Triticum), Pyricularia oryzae (as Triticum), Sclerophthora macrospora (as Triticum), Sitodiplosis mosellana (as Triticum), Soil-borne wheat mosaic virus (as Triticum), Tilletia caries (as Triticum), Tilletia controversa (as Triticum), Tilletia indica, Tilletia indica (as Triticum), Urocystis agropyri (as Triticum), Xanthomonas translucens pv. translucens (as Triticum); Major host: Meloidogyne naasi

Output structure

```
head(taxonomy$long_table, 10) %>%
   flextable::flextable() %>%
   flextable::bg(bg = "white", part = "all")
```

codeid	eppocode	prefname	level
57,021	DROSSU	Animalia	1
57,200	DROSSU	Arthropoda	2
77,963	DROSSU	Hexapoda	3
59,061	DROSSU	Insecta	4
58,226	DROSSU	Diptera	5
58,282	DROSSU	Drosophilidae	6
58,283	DROSSU	Drosophila	7
9,518	DROSSU	Drosophila suzukii	8
57,021	ERYTSZ	Animalia	1
57,200	ERYTSZ	Arthropoda	2

16

```
head(category$long_table, 10) %>%
    flextable::flextable() %>%
    flextable::bg(bg = "white", part = "all")
```

eppocode	nomcontinent	isocode	country	qlist	qlistlabel	yr_add	yr_del	yr_trans
DROSSU	Africa	MA	Morocco	X	Quarantine pest	2,018		
DROSSU	Africa	TN	Tunisia	X	Quarantine pest	2,012		
DROSSU	America	MX	Mexico	X	Quarantine pest	2,018		
DROSSU	Asia	JO	Jordan	1	A1 list	2,013		
DROSSU	Asia	KZ	Kazakhstan	1	A1 list	2,017		
DROSSU	Europe	TR	Turkey	1	A1 list	2,016		
DROSSU	RPPO/EU	9D	COSAVE	2	A2 list	2,018		
DROSSU	RPPO/EU	9M	EAEU	1	A1 list	2,016		
DROSSU	RPPO/EU	9A	EPPO	2	A2 list	2,011		
ERYTSZ								

```
write.csv(pests_names$preferred_names, "pest_pref_names.csv")
write.csv(hosts$long_table, "hosts_long.csv")
```

OR

```
saveRDS(pests_names, "pests.RDS")
saveRDS(hosts, "hosts.RDS")
```

18

That is it. That is all you need to know.

Or... is it...?

What are you going to do with your data?

Copy-paste to Spreadsheet

Process manually

Waste. Your. Time.

Make ugly charts with default settings

What can you achievie with spending just 15 more minutes on coding?

Simple functions

```r
library("tidyverse")
taxons_heatmap <- function(query, eppo_token) {
  query_names <- eppo_names_tables(query)

  hosts_eppocode <- query_names$pref_names %>%
    filter(grepl(paste0("^", query), fullname)) %>%
    .$eppocode

  query_pests <- eppo_tabletools_pests(token = eppo_token,
                  raw_eppocodes = hosts_eppocode, use_raw_codes = TRUE)

  unique_pests <- query_pests$long_table %>%
    filter(!is.na(pests_eppocode)) %>%
    .$pests_eppocode %>%
    unique

  pest_taxonomy <- eppo_tabletools_taxo(token = eppo_token,
                      raw_eppocodes = unique_pests,
                      use_raw_codes = TRUE)

  taxon_plot_df <- pest_taxonomy$compact_table %>%
    rename(pests_eppocode = eppocode) %>%
    left_join(query_pests$long_table) %>%
    group_by(eppocode) %>%
    count(taxonomy) %>%
    group_by(n_taxons = cut(n, breaks = c(seq(1,51,2), Inf), right = FALSE))

  return(taxon_plot_df)
}
```

24

It took ca. 14 seconds to execute.

```r
oryza_for_plot <- taxons_heatmap("Oryza", eppo_token)

oryza_for_plot %>%
  ggplot(aes(x = eppocode, y = taxonomy, fill = n_taxons)) +
  theme_bw() +
  geom_tile() +
  scale_fill_viridis_d(direction = -1) +
  theme(axis.text.x = element_text(angle = 90)) +
  xlab("Host eppocode")
```

It took ca. 11 secs seconds to execute

```r
cucumis_for_plot <- taxons_heatmap("Cucumis", eppo_token)

cucumis_for_plot %>%
  ggplot(aes(x = eppocode, y = taxonomy, fill = n_taxons)) +
  theme_bw() +
  geom_tile() +
  scale_fill_viridis_d(direction = -1, option = "C") +
  theme(axis.text.x = element_text(angle = 90), panel.grid = element_blank()) +
  xlab("Host eppocode")+
  labs(title = "Genus: Cucumis", subtitle = "Number of pest taxons", fill = "No. taxons") +
  guides(fill = guide_legend(reverse = TRUE))
```
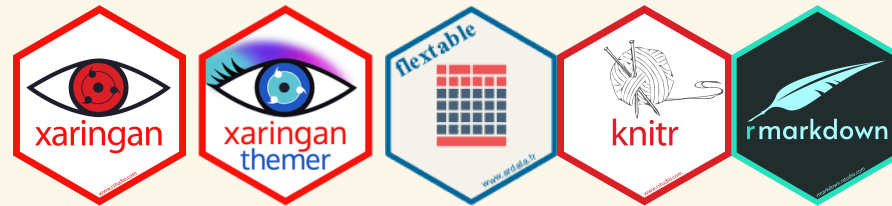
Genus: Cucumis
Number of pest taxons
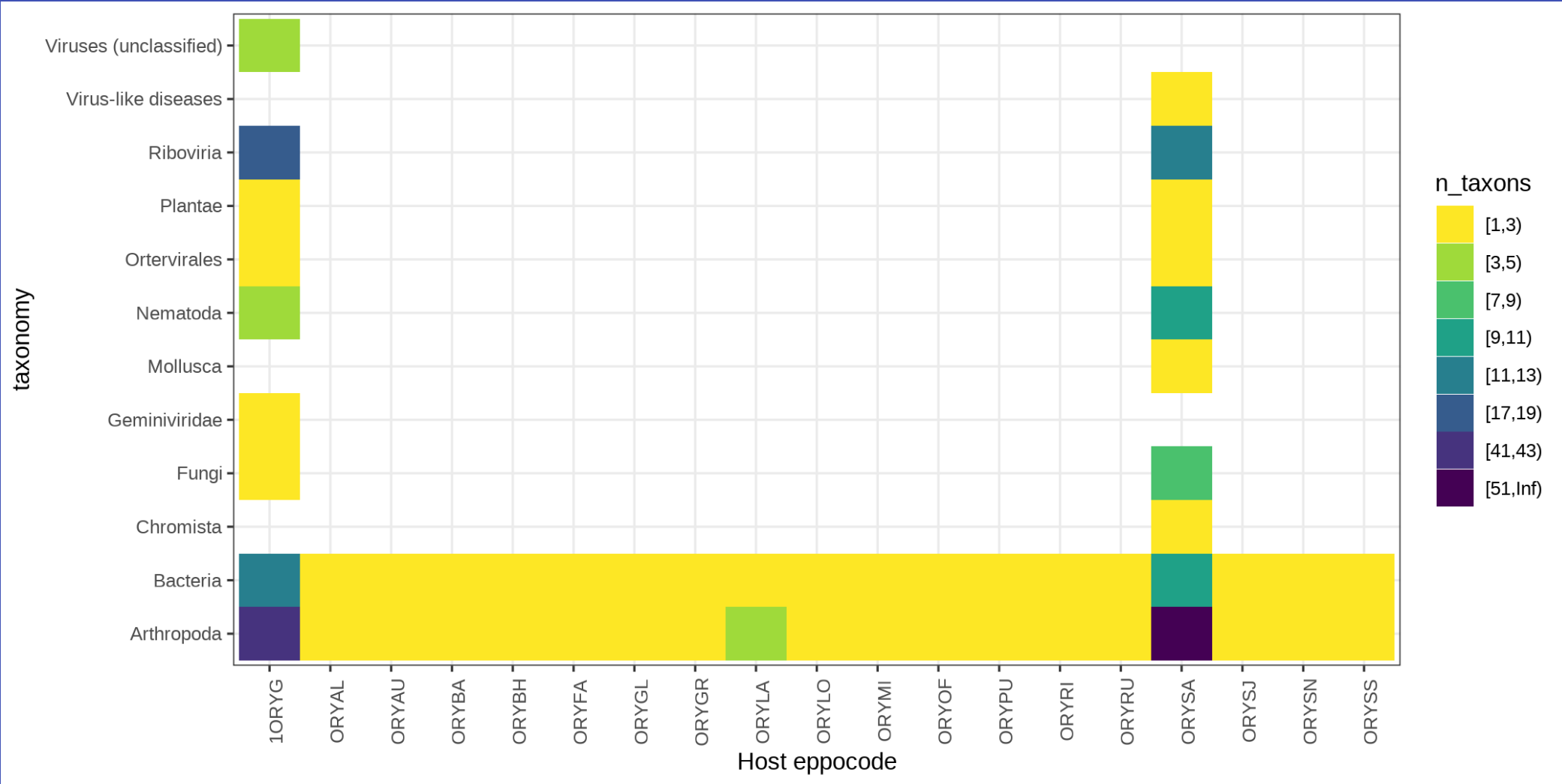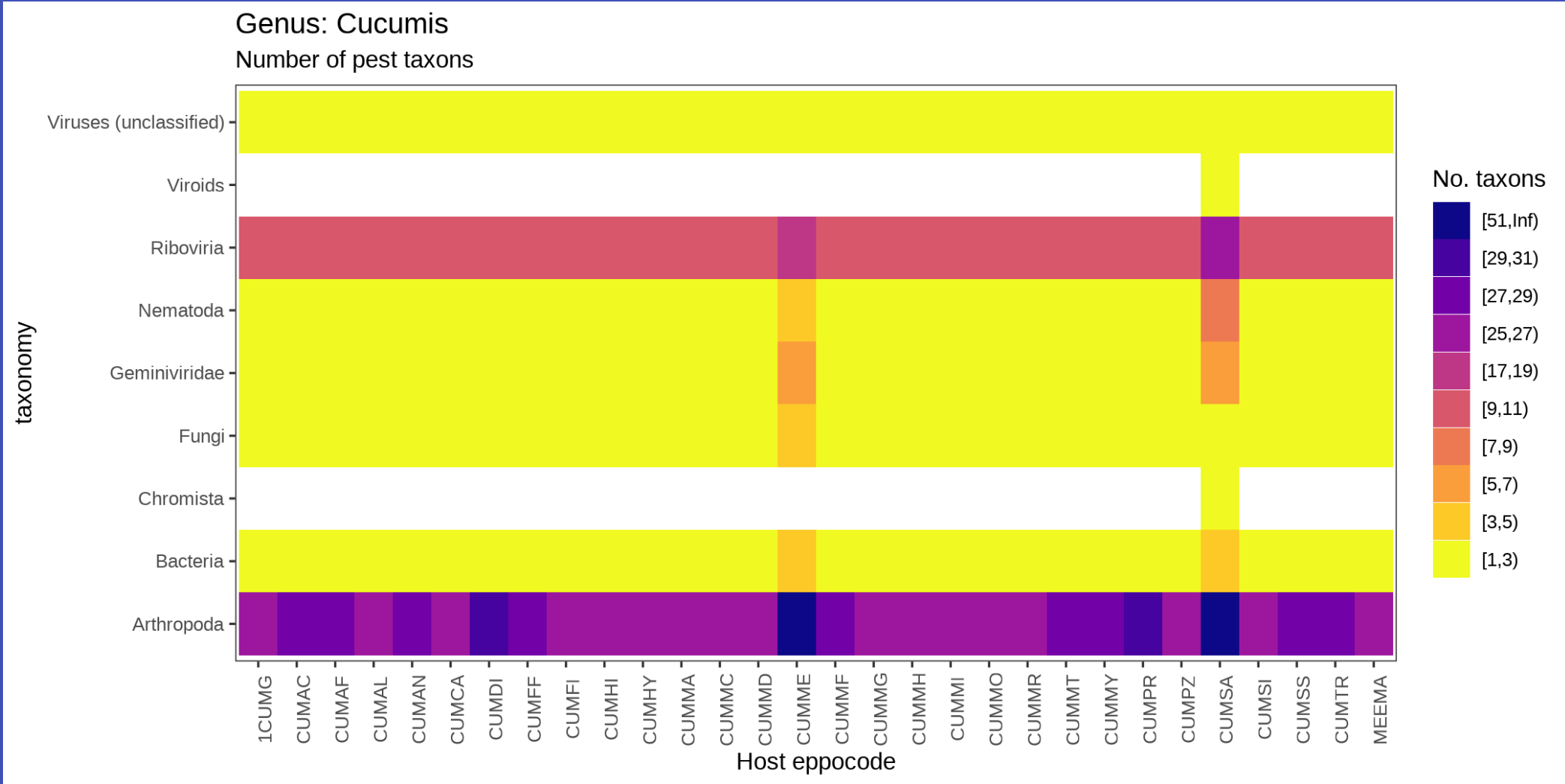
Use available solutions to automate your work.

Learn basic coding techniques and speed up things.

Stop wasting time on repetitive and boring tasks.

Thank you for listening!

Slides created via the R packages:

with some help of: