

Phylogenetics_SRV_viruses

Anton Zhelonkin

5/24/2022

Specifications

version

```
##  
## platform      _  
## arch          x86_64-pc-linux-gnu  
## os            linux-gnu  
## system        x86_64, linux-gnu  
## status  
## major         4  
## minor         2.0  
## year          2022  
## month         04  
## day           22  
## svn rev       82229  
## language      R  
## version.string R version 4.2.0 (2022-04-22)  
## nickname      Vigorous Calisthenics
```

Requirements

```
if (!("reutils" %in% installed.packages()))  
install.packages("reutils")  
library(reutils)  
sessionInfo()  
  
## R version 4.2.0 (2022-04-22)  
## Platform: x86_64-pc-linux-gnu (64-bit)  
## Running under: Ubuntu 20.04.4 LTS  
##  
## Matrix products: default  
## BLAS: /usr/lib/x86_64-linux-gnu/blas/libblas.so.3.9.0  
## LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.9.0  
##  
## locale:  
## [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C  
## [3] LC_TIME=ru_RU.UTF-8      LC_COLLATE=en_US.UTF-8  
## [5] LC_MONETARY=ru_RU.UTF-8  LC_MESSAGES=en_US.UTF-8  
## [7] LC_PAPER=ru_RU.UTF-8     LC_NAME=C  
## [9] LC_ADDRESS=C             LC_TELEPHONE=C  
## [11] LC_MEASUREMENT=ru_RU.UTF-8 LC_IDENTIFICATION=C  
##
```

```
## attached base packages:
## [1] stats      graphics  grDevices utils      datasets  methods   base
##
## other attached packages:
## [1] reutils_0.2.3
##
## loaded via a namespace (and not attached):
## [1] digest_0.6.29  bitops_1.0-7    magrittr_2.0.3  evaluate_0.15
## [5] rlang_1.0.2    stringi_1.7.6   cli_3.3.0       rstudioapi_0.13
## [9] rmarkdown_2.14 tools_4.2.0     stringr_1.4.0   RCurl_1.98-1.6
## [13] xfun_0.31      yaml_2.3.5      fastmap_1.1.0   compiler_4.2.0
## [17] htmltools_0.5.2 knitr_1.39
```

Description

This is a replication analysis of a paper by Zao et al. “A novel simian retrovirus subtype discovered in cynomolgus monkeys (*Macaca fascicularis*)”. The plan is simple: * download sequences by assesion numbers from GenBank (db-nucleotide) * align sequences of the newly discovered viral strain SRV * draw a phylogenetic tree comparing the new strain with the previously known strains

1. Data preparation

Preparing for data extraction

```
options(reutils.email = "my_email@email.com") # i`m a real person
```

find UIDs

```
# newly discovered SRV viral whole genomes search by assesion number given in the article
srv_new1 <- esearch(db = "nucleotide", term = "KU605777")
srv_new2 <- esearch(db = "nucleotide", term = "KU605778")
srv_new3 <- esearch(db = "nucleotide", term = "KU605779")

## Warning: CurlError: Resolving timed out after 10000 milliseconds
srv_new4 <- esearch(db = "nucleotide", term = "KU605780")

# previously known SRV viruses
srv1 <- esearch(db = "nucleotide", term = "M11841")
srv2 <- esearch(db = "nucleotide", term = "AF126467")
srv3 <- esearch(db = "nucleotide", term = "M12349")
srv4 <- esearch(db = "nucleotide", term = "FJ971077")
srv5 <- esearch(db = "nucleotide", term = "AB611707")
srv6_env <- esearch(db = "nucleotide", term = "AY598468")
srv7_pol <- esearch(db = "nucleotide", term = "AY594212")
serv <- esearch(db = "nucleotide", term = "U85505")

# all UIDs vector
uid <- c(srv_new1[1], srv_new2[1], srv_new3[1], srv_new4[1],
        srv1[1], srv2[1], srv3[1], srv4[1], srv5[1], srv6_env[1], srv7_pol[1], serv[1])
```

extract FASTA sequences by uids

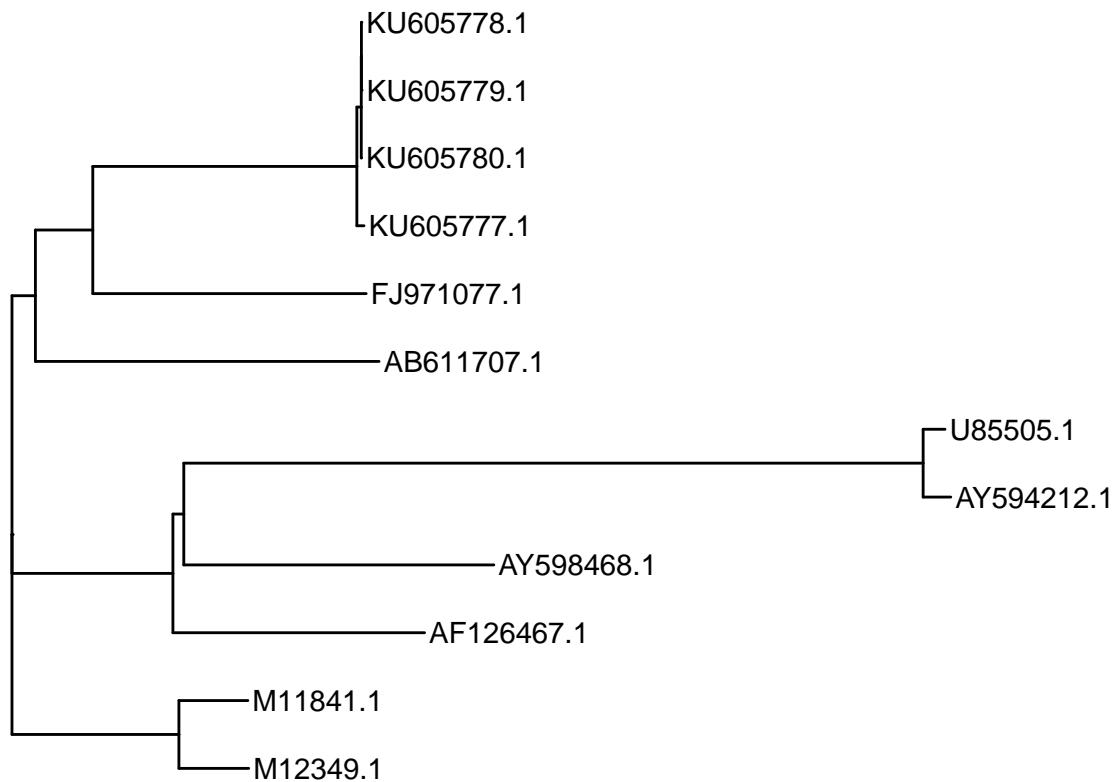
```
srv_new <- efetch(uid[1:4], db = "nucleotide", rettype = "fasta", retmode = "text")  
srv_seq_f <- efetch(uid[1:12], db = "nucleotide", rettype = "fasta", retmode = "text")
```

write FASTA file with new SRV genomes and all known SRV viruses with the newly discovered ones

```
write(content(srv_new), "srv_new_sequences")  
write(content(srv_seq_f), "srv_all_sequences")
```

Drawing tree based on modeltest-raxml pipeline

```
if (!("ggtree" %in% installed.packages()))  
  BiocManager::install("ggtree")  
  
library(ggtree)  
  
## ggtree v3.4.0 For help: https://yulab-smu.top/treedata-book/  
##  
## If you use the ggtree package suite in published research, please cite  
## the appropriate paper(s):  
##  
## Guangchuang Yu, David Smith, Huachen Zhu, Yi Guan, Tommy Tsan-Yuk Lam.  
## ggtree: an R package for visualization and annotation of phylogenetic  
## trees with their covariates and other associated data. Methods in  
## Ecology and Evolution. 2017, 8(1):28-36. doi:10.1111/2041-210X.12628  
##  
## Guangchuang Yu. Using ggtree to visualize data on tree-like structures.  
## Current Protocols in Bioinformatics. 2020, 69:e96. doi:10.1002/cpbi.96  
##  
## S Xu, Z Dai, P Guo, X Fu, S Liu, L Zhou, W Tang, T Feng, M Chen, L  
## Zhan, T Wu, E Hu, Y Jiang, X Bo, G Yu. ggtreeExtra: Compact  
## visualization of richly annotated phylogenetic data. Molecular Biology  
## and Evolution. 2021, 38(9):4039-4042. doi: 10.1093/molbev/msab166  
  
srv_tr <- read.tree("srv_all_raxml.raxml.bestTree")  
ggtree(srv_tr) + geom_tiplab() + xlim(0,0.8)
```



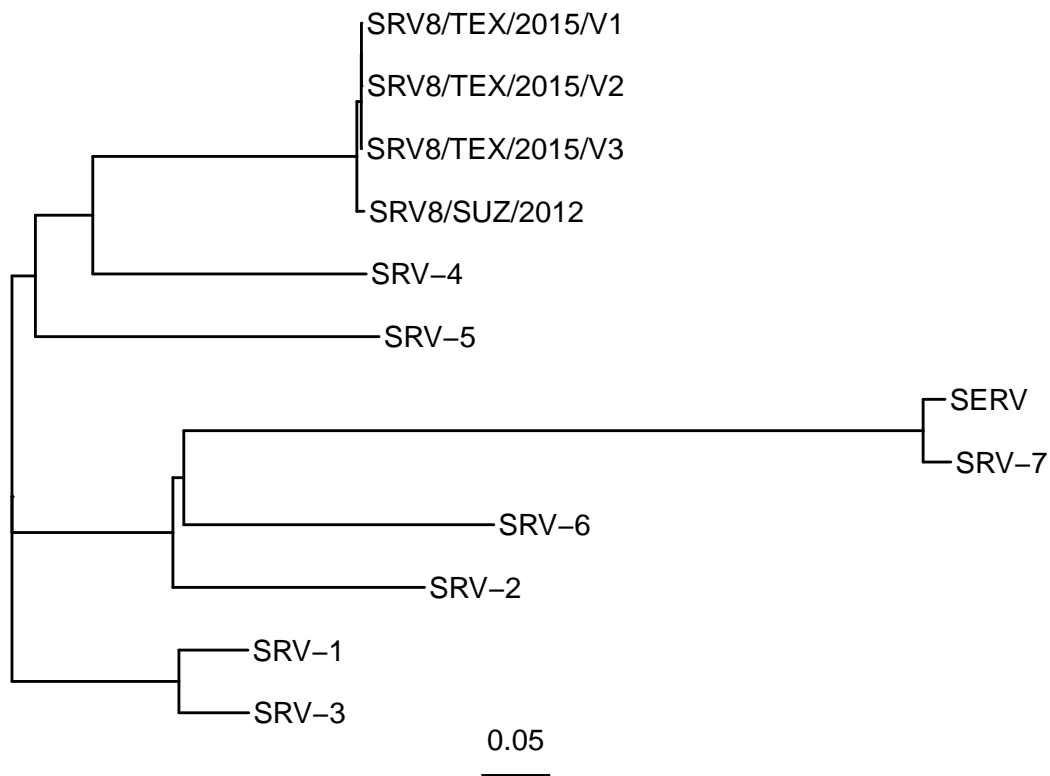
sabing tree

```
png("srv_all_raxml_bestTree.png")
ggtree(srv_tr) + geom_tiplab() + xlim(0,0.8)
dev.off()
```

```
## pdf
## 2
```

Renaming tip labels in tree

```
srv_tr_ren <- srv_tr
srv_tr_ren$tip.label <- c("SRV-5", "SRV-4", "SRV8/TEX/2015/V2", "SRV8/TEX/2015/V1",
                          "SRV8/TEX/2015/V3", "SRV8/SUZ/2012", "SRV-3", "SRV-1", "SRV-7",
                          "SERV", "SRV-6", "SRV-2")
ggtree(srv_tr_ren, ladderize=TRUE)+ geom_tiplab() + xlim(0,0.8) + geom_treescale(width = 0.05)
```



saving tree

```
png("srv_all_raxml_named_bestTree.png")
ggtree(srv_tr_ren, ladderize=TRUE)+ geom_tiplab() + xlim(0,0.8) + geom_treescale(width = 0.05)
dev.off()
```

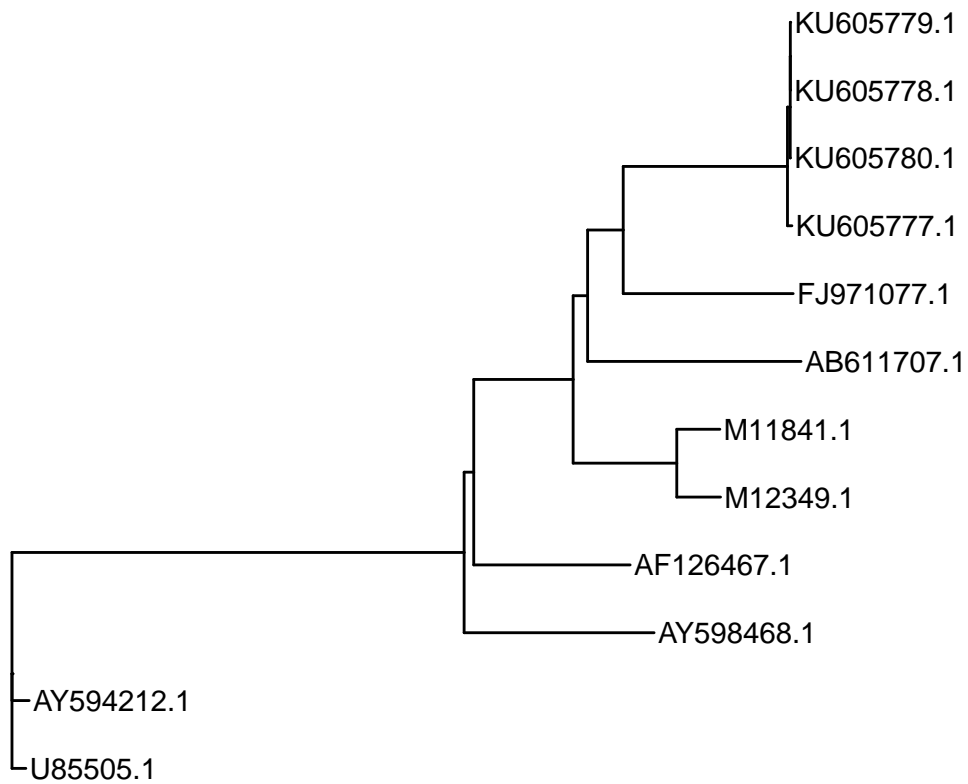
```
## pdf
## 2
```

Drawing tree based on PhyML3 Bootstrap 100

Analysis run at http://phylogeny.lirmm.fr/phylo.cgi/one_task.cgi?task_type=phym1 * Substitution model: GTR

* Bootstrapped data sets: 100

```
phy_tree <- read.tree("phym13_all_tree.nwk")
ggtree(phy_tree) + geom_tiplab() + xlim(0,2)
```

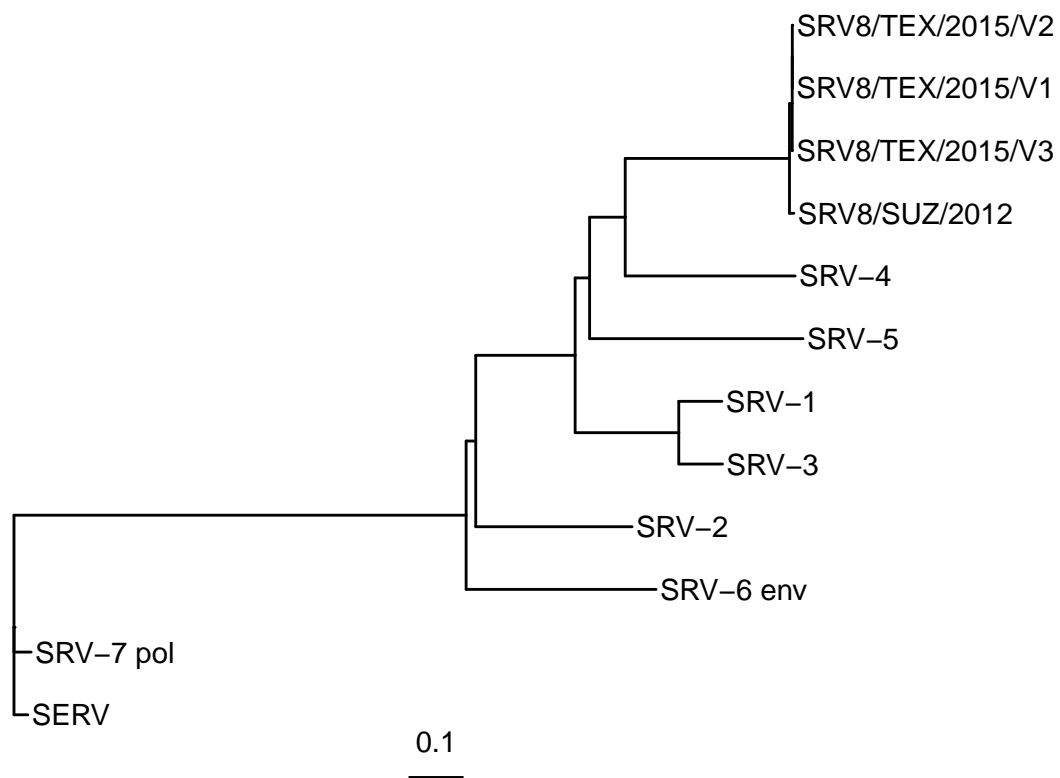


```
### Renaming tip labels in tree
```

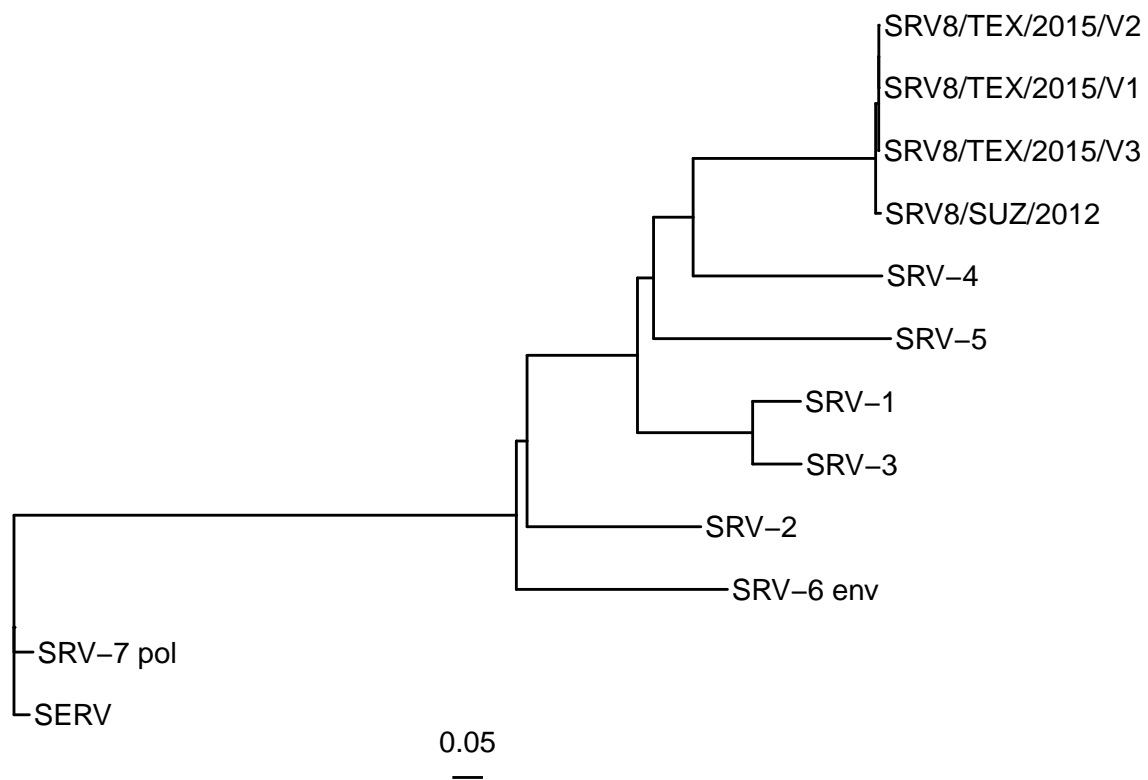
```
phy_tree_ren <- phy_tree
phy_tree_ren$tip.label <- c("SRV-4", "SRV8/SUZ/2012", "SRV8/TEX/2015/V1",
                             "SRV8/TEX/2015/V2", "SRV8/TEX/2015/V3", "SRV-5",
                             "SRV-3", "SRV-1", "SRV-2", "SRV-6 env", "SERV",
                             "SRV-7 pol")
ggtree(phy_tree_ren, ladderize=TRUE, scale = 0.5)+ geom_tiplab() + xlim(0,2) + geom_treescale()
```

```
## Warning: Ignoring unknown parameters: scale
```

```
## Ignoring unknown parameters: scale
```



```
ggtree(phy_tree_ren, ladderize=TRUE)+ geom_tiplab() + xlim(0,1.8) +  
  geom_treescale(width = 0.05)
```



saving tree

```
png("phym13_all_tree.png")
ggtree(phy_tree_ren, ladderize=TRUE)+ geom_tiplab() + xlim(0,1.8) +
  geom_treescale(width = 0.05)
dev.off()
```

```
## pdf
## 2
```

```
png("phym13_all_tree_res.png", width = 1200, height = 600)
ggtree(phy_tree_ren, ladderize=TRUE, size = 1.2)+ geom_tiplab(size = 10) + xlim(0,1.8) +
  geom_treescale(width = 0.05, fontsize = 10, linesize = 1.0)
dev.off()
```

```
## pdf
## 2
```