

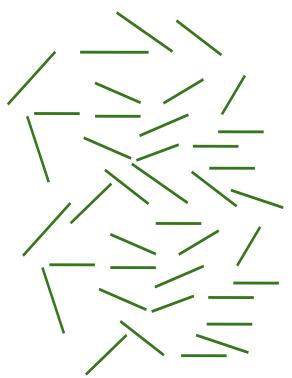
Virus typing tool and Data Visualization

Marta Giovanetti

giovanetti.marta@gmail.com

- **Virus typing tool**

Illumina/Ion Method



Quality control/pre-processing
FastQC and Trimmomatic

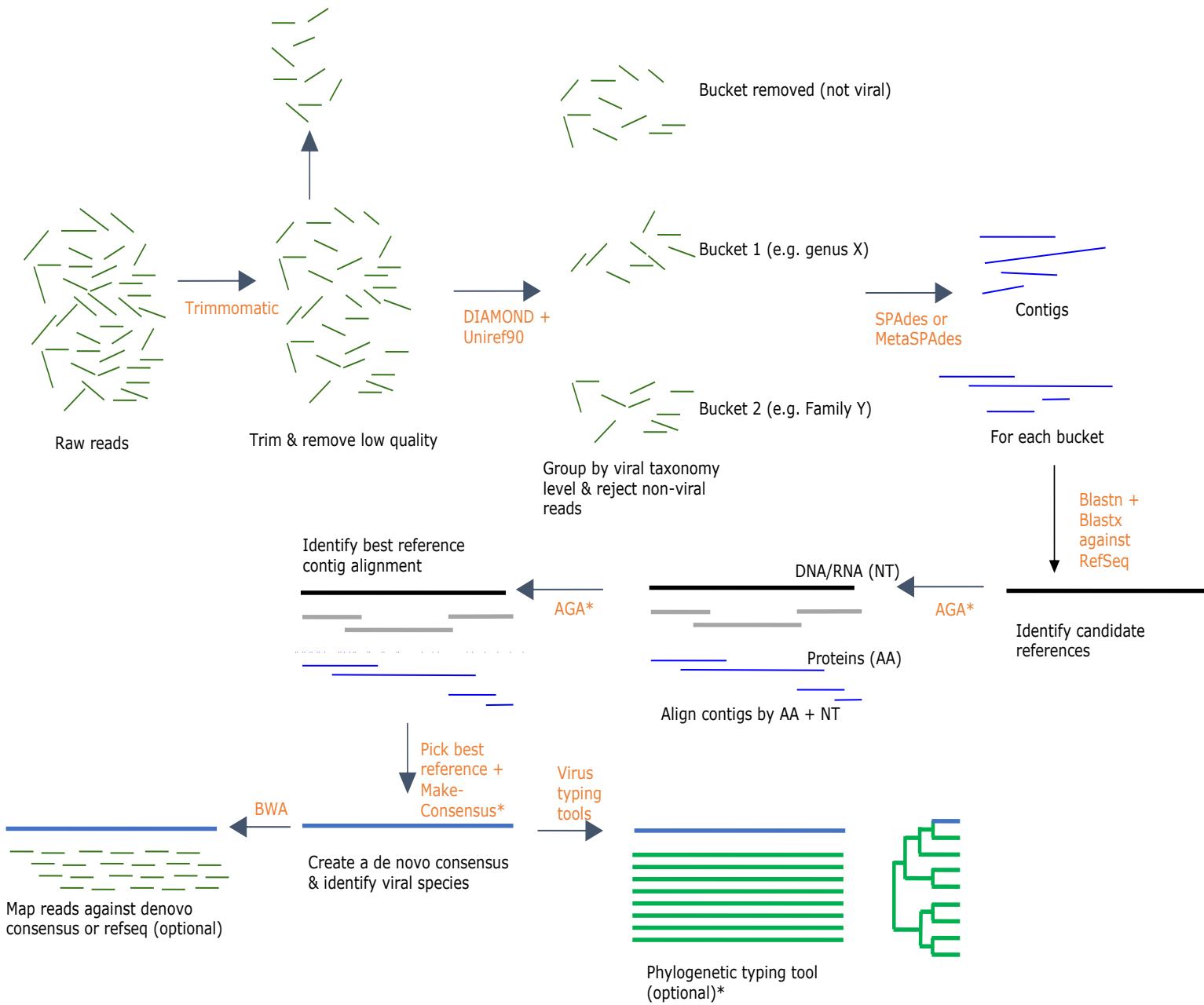
Filter and sorting
DIAMOND and UniRef90

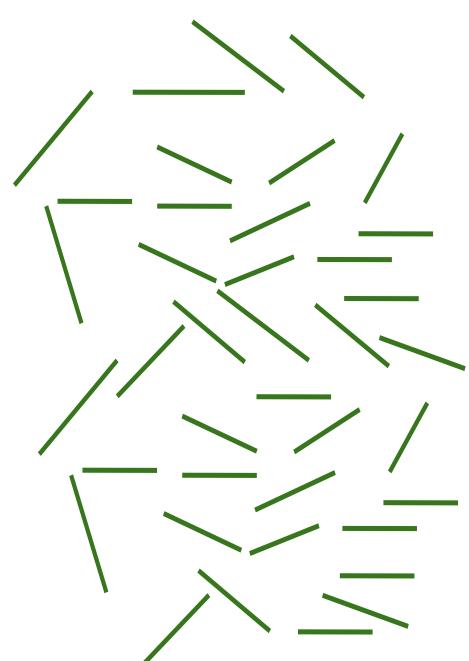
Assembly
SPAdes and MetaSPAdes

Search
AGA

Post-processing
Phylogenetic suite of tools



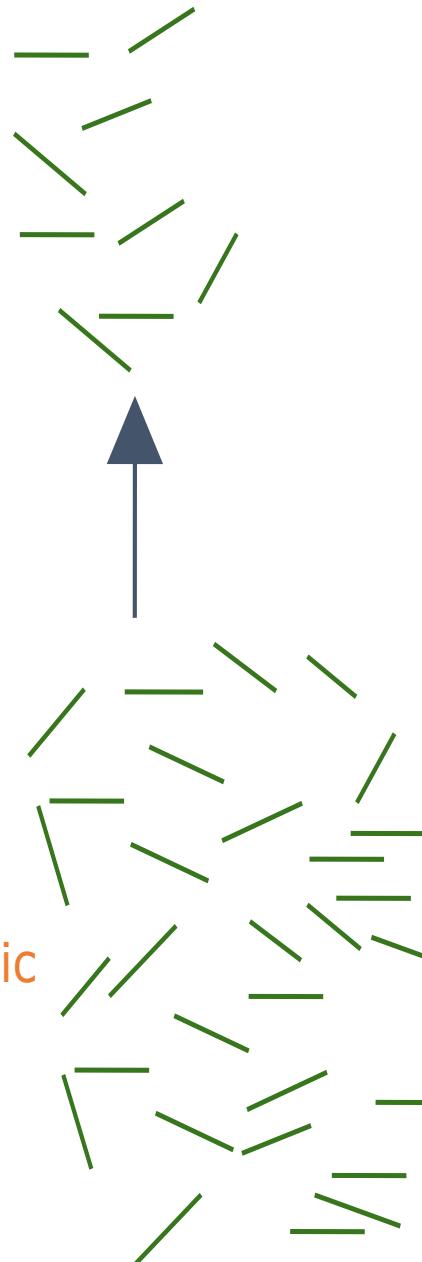




Raw reads

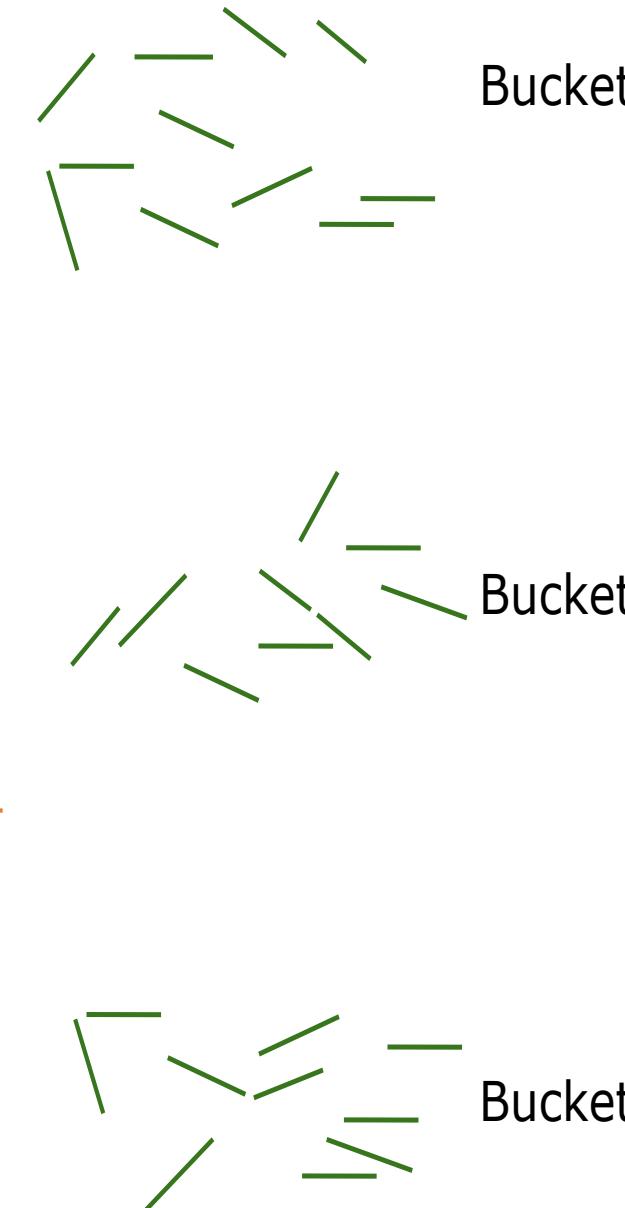


Trimmomatic



Trim & remove low quality

DIAMOND +
Uniref90



Bucket 1

Bucket 2

Bucket 3

[SUBMIT JOB](#)[MONITOR JOB](#)[DOCUMENTATION](#) ▾[PREMIUM](#)[LOGIN](#)

NO JOBS IN QUEUE

Genome Detective

GENOME DETECTIVE VIRUS TOOL

Version 2.43

How Genome Detective assigns SARS-CoV-2 (the virus that causes COVID-19 disease)[Read more](#)

GENOME DETECTIVE VIRUS TOOL VERSION 2.43

The Virus tool assigns taxonomic names to sequences from eukaryotic viruses and phages. The taxonomic rank which gets assigned is based on availability of reference genomes in RefSeq, and is mostly at species level. At the moment 11069 distinct taxonomic names are assigned based on 14498 reference sequences. Subtyping tools, for the identification of subspecies, are available for 19 viruses.

You can submit sequences/contigs or short reads NGS data. Up to 2000 FASTA sequences may be submitted at once. The free service only allows one NGS analysis at a time. For batch analyses of NGS data, please [contact us](#).

CHOOSE TYPE OF INPUT FILE

WHAT KIND OF INPUT FILE DO YOU HAVE?

- NGS short reads (FASTQ), e.g. Illumina, IonTorrent, ...
- 3GS long reads (FASTQ), e.g. Oxford Nanopore, PacBio, ...
- Contigs to be assembled (FASTA)
- Consensus sequences (FASTA)

INPUT

Submit one or more FASTA sequences to be typed individually. If you have raw NGS reads (short reads or long reads), please use the [Genome Detective Virus Tool](#) to assemble first. Subtyping tools will be linked in the results.

[Click here](#) to load some sample data.

Sequence

CLICK OR DROP FILE

(Paste or upload a FASTA sequence)

How Genome Detective assigns SARS-CoV-2 (the virus that causes COVID-19 disease)

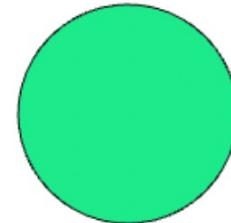
[Read more](#)

RESULTS

You may bookmark this page to revisit results of this job ([1de0f2e5-c09f-45a4-b5ed-7c63ef0b5316](#)) later.

SUMMARY

Blast/aga assignment	Sequences count	Percentage	Legend
Dengue virus type 2	1	100%	
Total	1	100%	



Name	Length	Report	Species	NT / AA Identity (%)	Genome
query2 2021-03-22	9448	Report	Dengue virus type 2	92.7572 97.3091	

Download results: [XML File](#) [Table \(Excel format\)](#) [Table \(CSV format\)](#) [Sequences \(Fasta format\)](#)

SEQUENCE ASSIGNMENT

Name query2|2021-03-22

Length 9448

ASSIGNMENT

Type Dengue virus type 2 (Taxonomy ID: [11060](#))

 [Continue to the dengue typing tool](#) to get a more specific assignment.

Reference Genome [NC_001474.2](#) (Length: 10723bp)

Host(s) Erythrocebus patas / Homo sapiens / Aedes aegypti / Aedes taylori / Aedes furcifer ([host info](#))

NT Identity (%) 92.7572

AA Identity (%) 97.3091

Number of stop codons 0

Number of CDS 1

ALIGNMENT

Alignment score 15278 (NT) + 20220 (AA) = 35498

Concordance (%) 92.3249

Alignment method Global, seeded, nucleotide + amino acids (AGA)

NT Alignment [Download alignment \(FASTA\)](#)

CDS Alignments [Download CDS alignments \(FASTA\)](#)

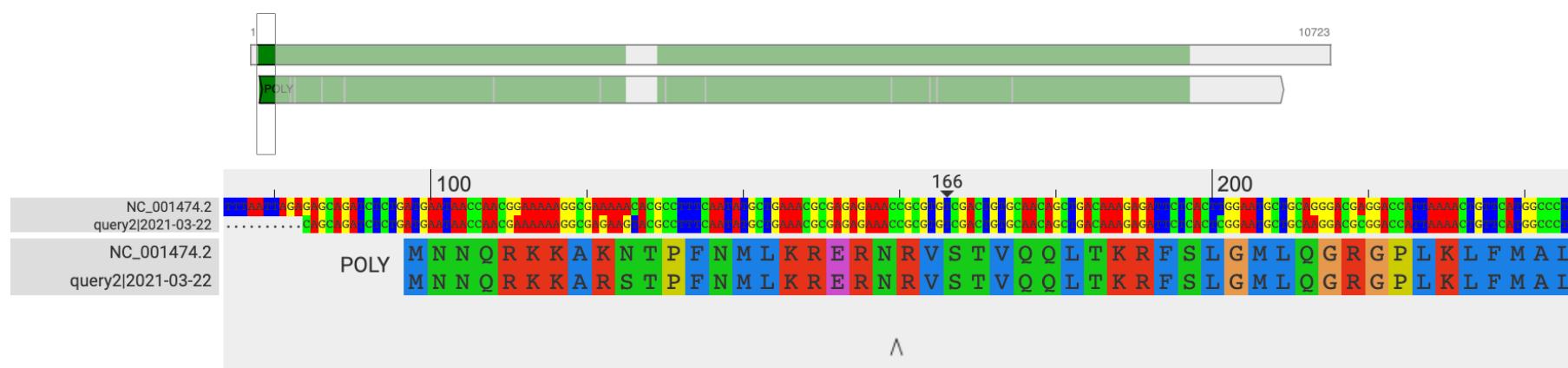
PHYLOGENETIC ANALYSIS

 [Continue to the dengue typing tool](#) to get a more specific assignment.

ALIGNMENT DETAILS

ALIGNMENT

Using NC_001474.2 (Dengue virus type 2 (taxon:11060)) as reference for alignment, numbering and genome annotations.



GENETIC DIVERSITY ANALYSIS

[SHOW MUTATIONS](#)

	Begin	End	Coverage	Score	Concordance	Matches	Identities	I/D/M/F*	Stop Codons
NT	84	9336	83.3%	15278	85.5%	8933 (100%)	8286 (92.8%)	0/0	
CDS									
POLY	1	3080	87.6%	20220	98.1%	2973 (100%)	2893 (97.3%)	0/0/0/0	0

SEQUENCE ASSIGNMENT

Name query2|2021-03-22

Length 9448

ASSIGNMENT

Type Dengue virus type 2 (Taxonomy ID: [11060](#))

 [Continue to the dengue typing tool](#) to get a more specific assignment.

Reference Genome [NC_001474.2](#) (Length: 10723bp)

Host(s) Erythrocebus patas / Homo sapiens / Aedes aegypti / Aedes taylori / Aedes furcifer ([host info](#))

NT Identity (%) 92.7572

AA Identity (%) 97.3091

Number of stop codons 0

Number of CDS 1

ALIGNMENT

Alignment score 15278 (NT) + 20220 (AA) = 35498

Concordance (%) 92.3249

Alignment method Global, seeded, nucleotide + amino acids (AGA)

NT Alignment [Download alignment \(FASTA\)](#)

CDS Alignments [Download CDS alignments \(FASTA\)](#)

PHYLOGENETIC ANALYSIS

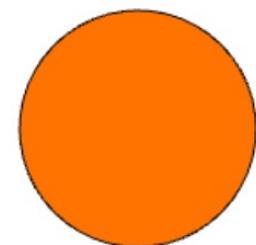
 [Continue to the dengue typing tool](#) to get a more specific assignment.

RESULTS

You may bookmark this page to revisit results of this job ([fa970c41-fbe0-441d-ba54-9fc41b761ef5](#)) later.

SUMMARY

Virus assignment	Genotype assignment	Sub-clustering	Sequences count	Percentage	Legend
Dengue virus type 2	DENV-2 Genotype III - Southern Asian-American	Not assigned	1	100%	
Total			1	100%	



Name	Length	Blast/Serotype	Genotype	Report	Genome
query2 2021-03-22	9448	Dengue virus type 2	DENV-2 Genotype III - Southern Asian-American	Report	

Download results: [Table \(Excel format\)](#) [Table \(CSV format\)](#) [Sequences \(Fasta format\)](#)

PHYLOGENETIC ANALYSIS DETAILS (SEROTYPE)

- Assignment: DENV-2 Genotype III - Southern Asian-American
- Bootstrap support: 100.0, bootstrap inside 100.0, bootstrap outside 0.0
- Download the alignment ([NEXUS format](#), [FASTA format](#))
- Phylogenetic Tree (export as [PDF](#), [NEXUS Format](#))

TREE CONTROLS

Layout

Rectilinear 

Transform

None 

Show labels

Highlight clusters

Color branches

DENV-2 Genotype I -

American

DENV-2 Genotype II -

Cosmopolitan

DENV-2 Genotype III -

Southern Asian-

American

DENV-2 Genotype IV -

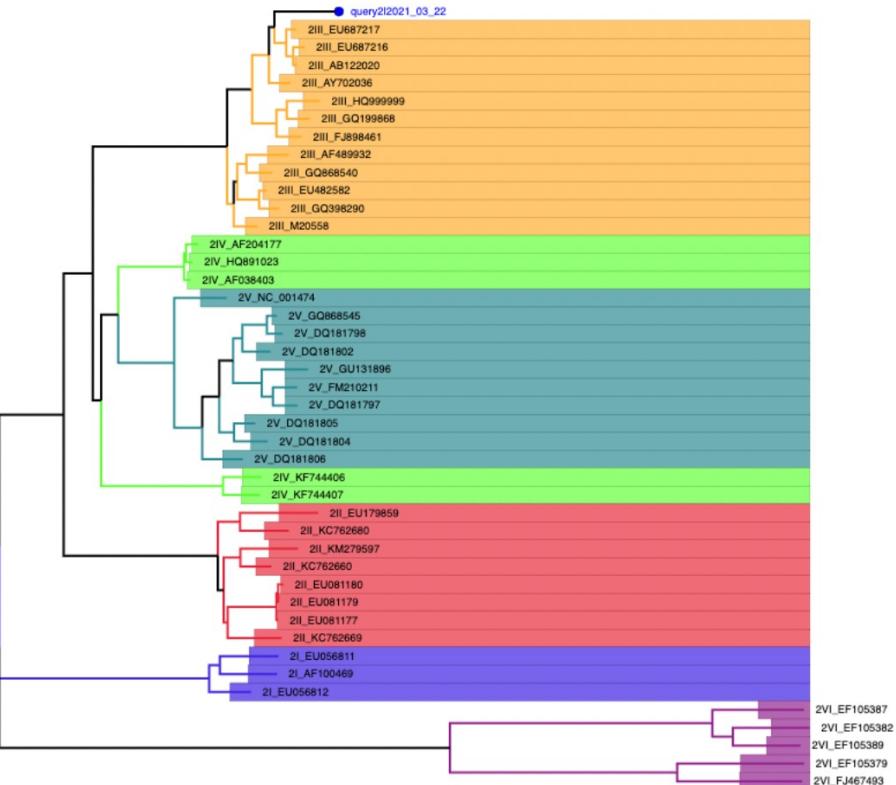
Asian II

DENV-2 Genotype V -

Asian I

DENV-2 Genotype VI -

Sylvatic



- View the [PAUP* Log file](#) (Contains bootstrap values)



CORONAVIRUS TYPING TOOL

CORONAVIRUS TYPING TOOL

Version 1.9

How Genome Detective assigns 2019-nCoV

[Read more](#)

CORONAVIRUS TYPING TOOL

This tool is designed to use Blast and phylogenetic methods in order to identify the Coronavirus types and genotypes of a nucleotide sequence.

The Coronavirus typing tool also includes a Wuhan Coronavirus genome, this sequence was generously shared by Professor Yong-Zhen Zhang and colleagues via [this post on virological.org](#). Professor Yong-Zhen Zhang and colleagues ask that you communicate with them if you wish to publish results that use the sequence that they share in a journal. We gratefully acknowledge their contribution.

Note for batch analysis: The tool accepts up to 2000 sequences at a time.

INPUT

Submit one or more FASTA sequences to be typed individually. If you have raw NGS reads (short reads or long reads), please use the [Genome Detective Virus Tool](#) to assemble first. Subtyping tools will be linked in the results.
[Click here](#) to load some sample data.

Sequence

CLICK OR DROP FILE

(Paste or upload a FASTA sequence)

- **Data Visualization**

Online sources to visualize your phylogeny

FigTree v.1.4.4	tree.bio.ed.ac.uk/software/figtree
MicroReact	microreact.org
ggtree (R)	https://doi.org/10.1111/2041-210X.12628
ape (R)	http://ape-package.ird.fr/
Ete2 (Python)	http://etetoolkit.org/
Dendroscope	http://dendroscope.org/
PhyloGeoTool	https://github.com/rega-cev/phylogeotool
Baltic (Python)	https://github.com/evogytis/baltic
GraPhlAn	https://bitbucket.org/nsegata/graphlan/wiki/Home
iTOL	https://itol.embl.de/
IcyTree	https://icytree.org/
PastML	https://pastml.pasteur.fr/
DensiTree	https://www.cs.auckland.ac.nz/~remco/DensiTree/
Seraphim	https://github.com/sdellicour/seraphim
Nextstrain	https://nextstrain.org/
SpreadD3	https://rega.kuleuven.be/cev/ecv/software/SpreeD3
Other useful resources:	colorbrewer2 colorbrewer2.org Batch geocoding www.storybench.org/geocode-csv-addresses-r/

Online resources for mapping your data

Google Earth	https://www.google.com/earth/
MapBox Studio	https://docs.mapbox.com/
Tableau	https://www.tableau.com
SimpleMapper	https://www.simplemappr.net/
ggmap (R)	https://cran.r-project.org/
QGIS	www.qgis.org
arcGIS	http://www.esri.com/software/arcgis
ggmap (R)	https://cran.r-project.org/web/packages/ggmap/
R spatial analysis	https://cran.r-project.org/web/views/Spatial.html
matplotlib (Python)	matplotlib.org/
D3 (Javascript)	d3js.org/
Neatline	https://neatline.org/

Other useful
resources:

colorbrewer2	colorbrewer2.org
Batch geocoding	www.storybench.org/geocode-csv-addresses-r/

Make figures in a fancy way

Visualising trees using FigTree

- Graphical viewer for phylogenetic trees and a program for preparing *publication ready figures* (e.g. PDF, JPG, etc). Originally designed to display and summarize annotated trees from BEAST
- Reads nexus and newick tree files (and tab-delimited metadata file)
- Flexible display of posterior probabilities, height, rate, trait on each branch/node and time axis
- Will open directly files with extension .tree and .tre
- Available from <http://tree.bio.ed.ac.uk/software/figtree/>



Open the [west-african-ebola.tre](#)

File -> Import Annotations (choose: “[west-african-ebola.metadata.txt](#)”)

Tree -> Annotate Nodes from Tips (choose: “[country](#)”)

Visualising trees using Microreact

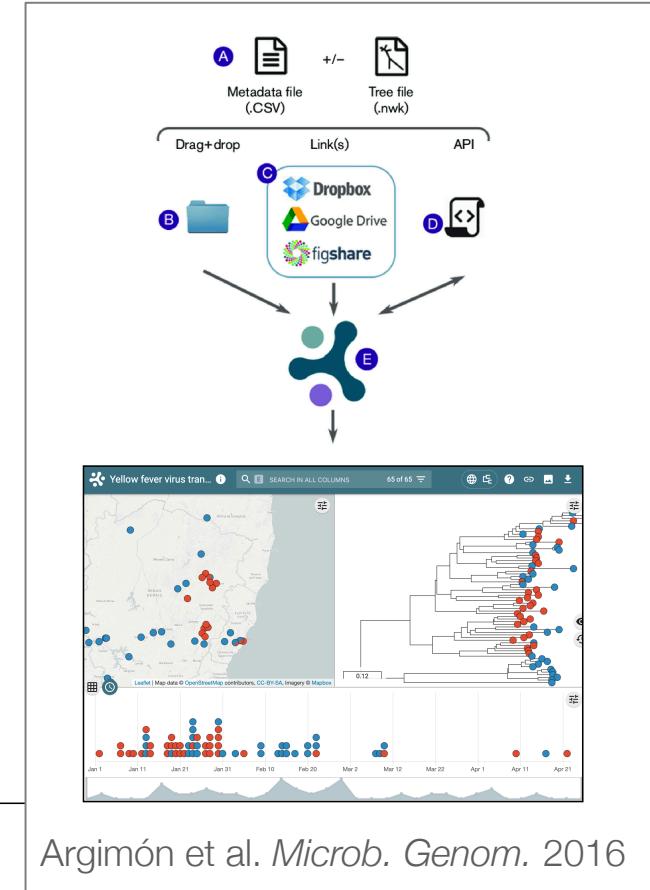
- Web interactive application for the easy visualization of datasets consisting of any combination of trees, spatial, temporal and associated metadata
- Reads a CSV-delimited metadata file (and a corresponding newick tree file)
- Allows users to upload, visualize and explore phylogenies linked to spatial locations
- Available at <https://microreact.org/>



Go to <https://microreact.org> (choose: UPLOAD)

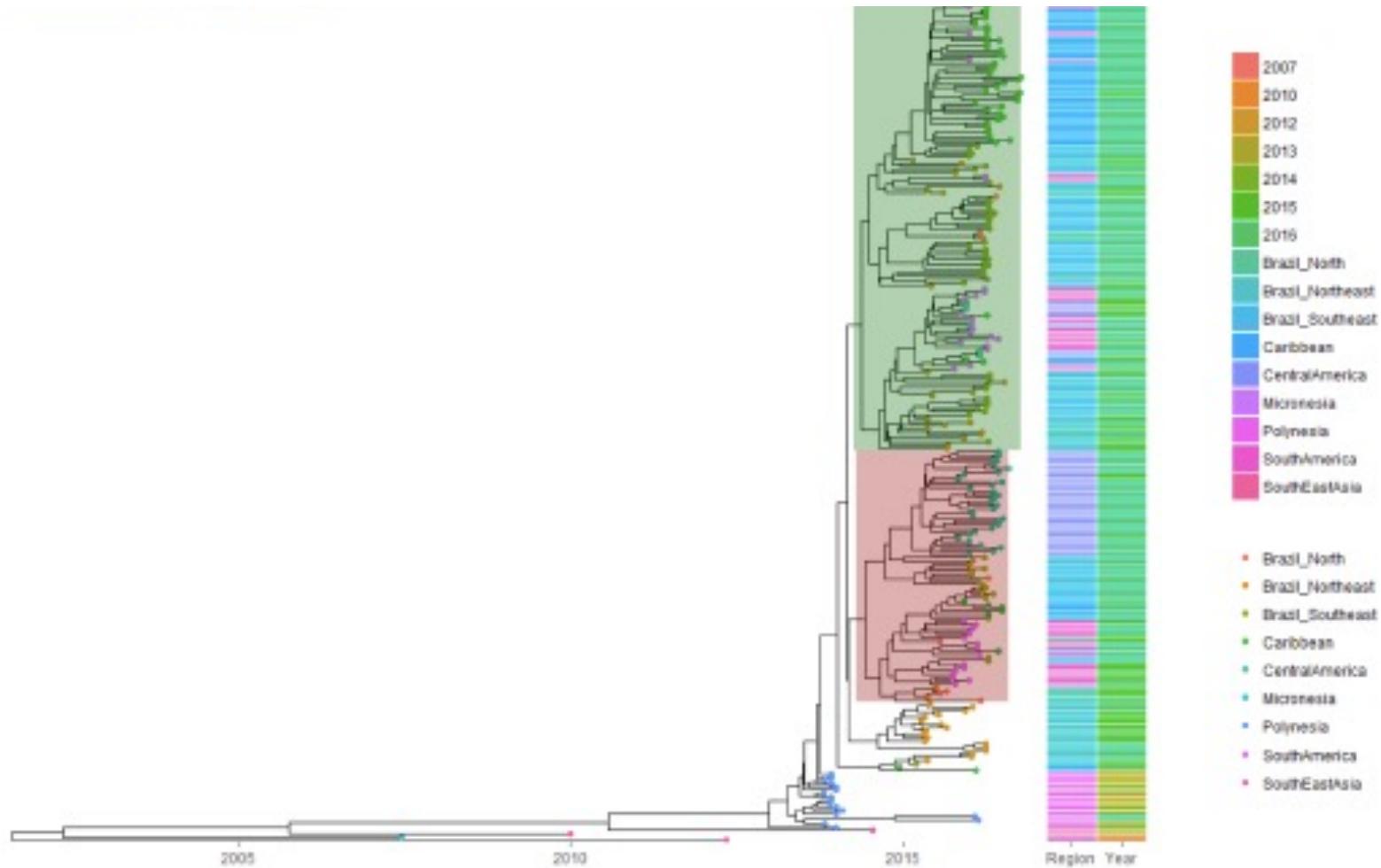
Drag and Drop nwk tree (choose: "YFV_outbreak.tre")

Drag and Drop CSV metadata (choose: "YFV_outbreak_metadata.csv")



Argimón et al. *Microb. Genom.* 2016

ggtree in R



Plotting roo-to-tip regression using R (ggplot)

```
if (!require("ggplot2")){
  install.packages("ggplot2")
}
library("ggplot2") #load the library
#remotes::install_github("wilkelab/cowplot")
library("cowplot")
theme_set(theme_cowplot())
install.packages("cowplot")

if (! require("lubridate")){
  install.packages("lubridate")
  library(lubridate)
}

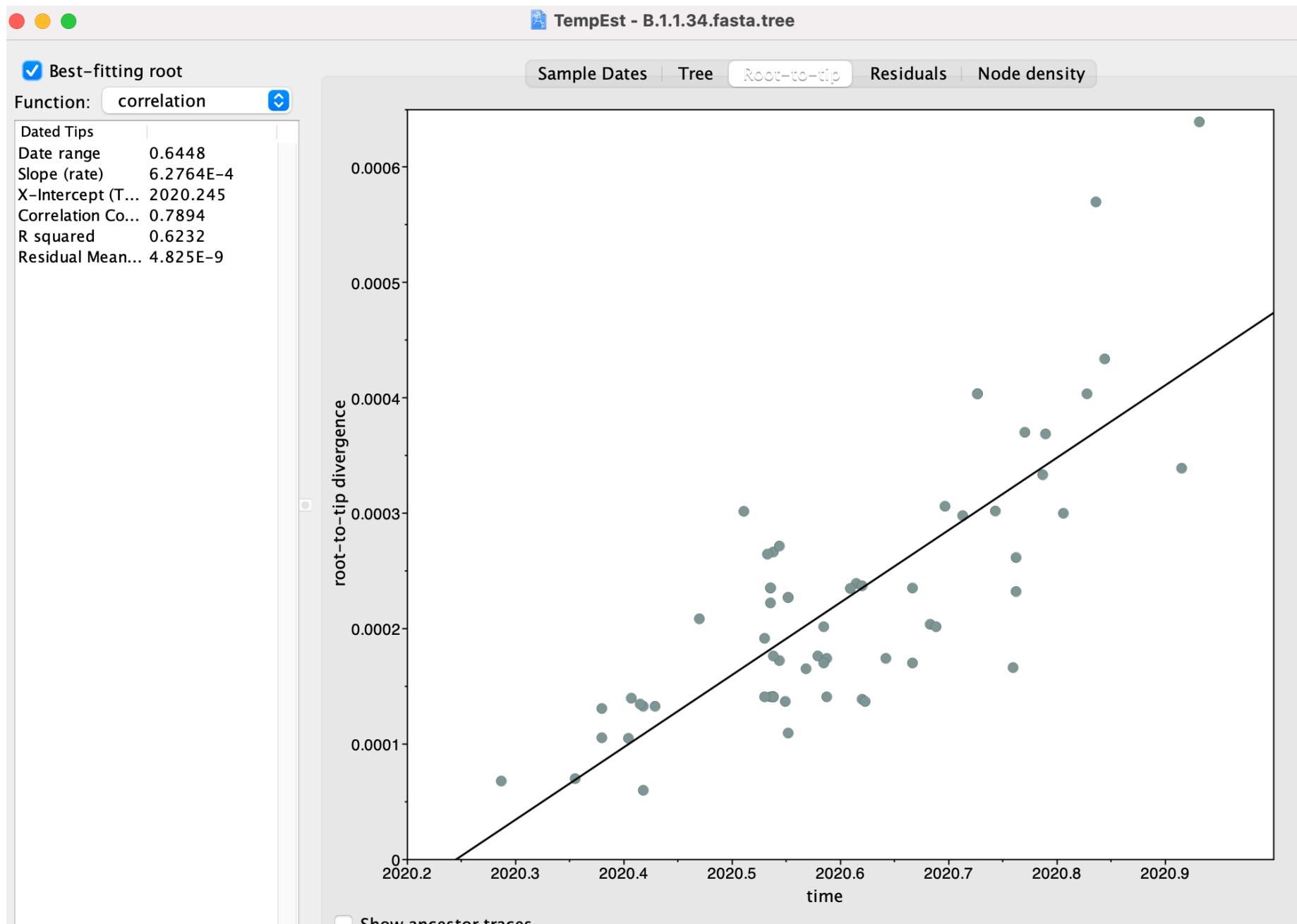
myFileData <- "/Users/Mittenavoig/Desktop/Marta_Gio/Practise_MG/Data_Visualization/GGPLOT\ -\ Tempest/Tempest/R
if (!(file.exists(myFileData))){
  stop("Error:directory does not exist, please check if the path is correct.")
}
#DATA
data = read.table(myFileData, header=T) # read from the file
```

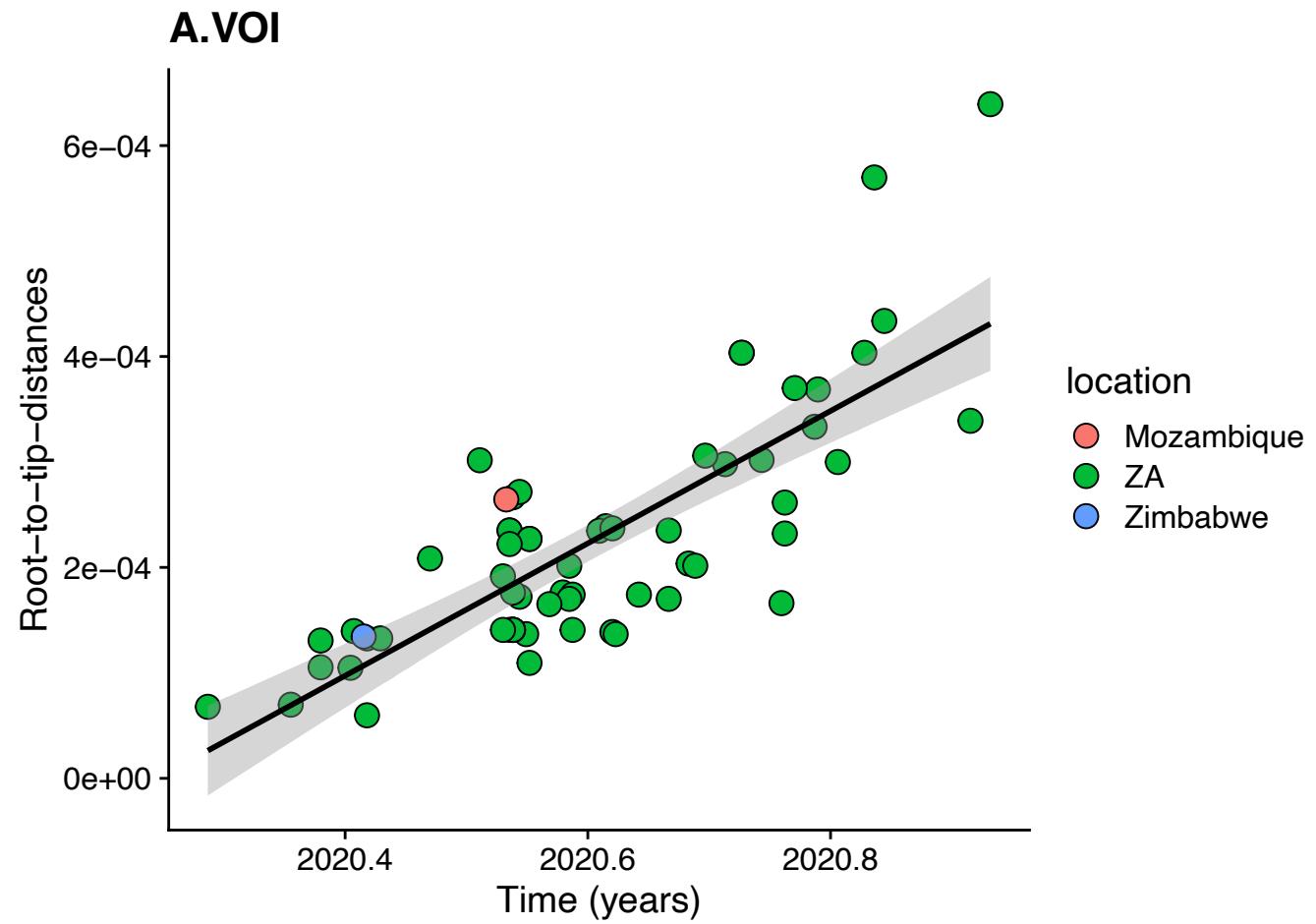
```
# GGPLOT
ggplot(data, aes(date,distance))+  
geom_point()+
geom_smooth(method=lm,se=T)+
#theme_bw()+
ylab("Distance")+xlab("Date")+ggtitle("Tempest")

#GGPLOT_colour
ggplot(data, aes(date, distance, location))+  
geom_point(aes(fill = location, shape=location), size = 4)+  
scale_shape_manual(values = c(21, 21, 21, 21, 21, 21, 21, 21,21, 21))+
geom_smooth(method=lm,se=T, colour="black")+
#theme_bw()+
ylab("Root-to-tip-distances")+
xlab("Time (years)")+
ggtitle("A.VOI")

#Additionally_Annotating using cc and r2
annotate("text", fontface =2,x=2021.05, y=6e-04, label="cofficient correletion=0.51", col='tomato3')
annotate("text", fontface =2,x=2021.2, y=5e-04, label="r2=0.24", col='tomato3')

#Converting decimal year to date format
library(lubridate)
x=c(2020.4, 2020.6,2020.8)
f <- format(date_decimal(x), "%d-%m-%Y")
f
```





Let's practice!