

# Updated introduction of Gamma in Uruguay

Timely vaccination and strengthening of genomic surveillance curbed consecutive waves of SARS-CoV-2 variants of concern despite evidence of cryptic transmission

## Procedure:

1. Read the annotated tree (region) and find Uruguay sequence clusters based on pattern changes in the “region” of the sample across the tree. Each region will have a unique cluster number.
2. Extract the patterns in a list format.
3. Convert list to data frame format and extract Uruguay cluster sequences. Each uruguayan sequence will have the cluster number detected in the tree.
4. Based on the uruguayan clusters (and VOC specific clusters), the MRCA of this sequences will be estimated using the MRCA function from phylobase.
5. Once the MRCA is found, the dates from the MRCA (tMRCA) will be retrieved using the node label from the dates.tsv file obtained after time-scaling the ML tree with TreeTime joint inference.
6. Using the information of the ancestral nodes in dates.tsv, find the location (outside Uruguay) of the ancestral node of the cluster using the annotated tree loaded with the read.beast() function.
7. If the location of the ancestral node is in Uruguay, use an iterative function that subtracts by 1 in the parent node column until Uruguay is no longer detected (use with caution).

```
library(ape)
library(treeio)
```

8. After the procedure a data frame containing the tMRCA (date & numeric\_date), the parent node outside Uruguay (parent) and the location of the ancestral node of the UY clusters (location) will be obtained.

```
## treeio v1.26.0 For help: https://yulab-smu.top/treedata-book/
##
## If you use the ggtree package suite in published research, please cite
```

```
## the appropriate paper(s):
##
## LG Wang, TTY Lam, S Xu, Z Dai, L Zhou, T Feng, P Guo, CW Dunn, BR
## Jones, T Bradley, H Zhu, Y Guan, Y Jiang, G Yu. treeio: an R package
## for phylogenetic tree input and output with richly annotated and
## associated data. Molecular Biology and Evolution. 2020, 37(2):599-603.
## doi: 10.1093/molbev/msz240
##
## G Yu. Data Integration, Manipulation and Visualization of Phylogenetic
## Trees (1st ed.). Chapman and Hall/CRC. 2022. ISBN: 9781032233574
##
## Guangchuang Yu. Using ggtree to visualize data on tree-like structures.
## Current Protocols in Bioinformatics. 2020, 69:e96. doi:10.1002/cpbi.96
```

```
library(ggtree)
```

```
## ggtree v3.10.1 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
##
## Shuangbin Xu, Lin Li, Xiao Luo, Meijun Chen, Wenli Tang, Li Zhan, Zehan
## Dai, Tommy T. Lam, Yi Guan, Guangchuang Yu. Ggtree: A serialized data
## object for visualization of a phylogenetic tree and annotation data.
## iMeta 2022, 1(4):e56. doi:10.1002/imt2.56
##
##
## Attaching package: 'ggtree'
##
## The following object is masked from 'package:ape':
##
##     rotate
```

```
library(ggplot2)
library(ape)
library(treeio)
library(phylostrate)
library(tidytree)
```

```
## If you use the ggtree package suite in published research, please cite
## the appropriate paper(s):
##
## Guangchuang Yu, Tommy Tsan-Yuk Lam, Huachen Zhu, Yi Guan. Two methods
## for mapping and visualizing associated data on phylogeny using ggtree.
```

```

## Molecular Biology and Evolution. 2018, 35(12):3041-3043.
## doi:10.1093/molbev/msy194
##
## G Yu. Data Integration, Manipulation and Visualization of Phylogenetic
## Trees (1st ed.). Chapman and Hall/CRC. 2022. ISBN: 9781032233574

##
## Attaching package: 'tidytree'

## The following object is masked from 'package:treeio':
##
##     getNodeNum

## The following objects are masked from 'package:ape':
##
##     drop.tip, keep.tip

## The following object is masked from 'package:stats':
##
##     filter

library(ggnewscale)
library(stringr)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:ape':
##
##     where

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(RColorBrewer)
library(phytools)

## Loading required package: maps

##
## Attaching package: 'phytools'

## The following object is masked from 'package:treeio':
##
##     read.newick

```

```
library(phylobase)
```

```
##
## Attaching package: 'phylobase'

## The following object is masked from 'package:phytools':
##
##   readNexus

## The following objects are masked from 'package:tidytree':
##
##   ancestor, MRCA

## The following object is masked from 'package:ggtree':
##
##   MRCA

## The following objects are masked from 'package:treeio':
##
##   ancestor, MRCA

## The following object is masked from 'package:ape':
##
##   edges
```

```
library(phylotate)
```

```
tre <- read.nexus("treetime_out_joint/fixed_mugration_region/annotated_tree.nexus")
meta <- read.csv("fx_final_metadata-dataset1.tsv",
  sep = "\t", header = T)
head(meta)
```

### Load annotated tree from dataset and metadata

```
##
##                                     strain      date
## 1  UY-CUY29-005480|EPI_ISL_NA|Uruguay|Rocha|AY.20|2021-09-30 2021-09-30
## 2  UY-CUY29-005481|EPI_ISL_NA|Uruguay|Rocha|AY.20|2021-09-30 2021-09-30
## 3  UY-CUY29-005487|EPI_ISL_NA|Uruguay|Colonia|AY.20|2021-09-30 2021-09-30
## 4  UY-CUY29-005488|EPI_ISL_NA|Uruguay|Canelones|AY.43|2021-09-30 2021-09-30
## 5  UY-CUY28-005443|EPI_ISL_NA|Uruguay|NA|AY.39|2021-09-29 2021-09-29
## 6  UY-CUY29-005486|EPI_ISL_NA|Uruguay|Florida|AY.25|2021-09-28 2021-09-28
##   accession location      region country      city subregion pango_lineage
## 1 EPI_ISL_NA  Uruguay South America Uruguay   Rocha    Rocha      AY.20
## 2 EPI_ISL_NA  Uruguay South America Uruguay   Rocha    Rocha      AY.20
## 3 EPI_ISL_NA  Uruguay South America Uruguay  Colonia  Colonia      AY.20
## 4 EPI_ISL_NA  Uruguay South America Uruguay Canelones Canelones      AY.43
## 5 EPI_ISL_NA  Uruguay South America Uruguay    <NA>   Uruguay      AY.39
## 6 EPI_ISL_NA  Uruguay South America Uruguay  Florida  Florida      AY.25
```

```
dim(meta)
```

```
## [1] 5173    9
```

```
meta <- meta[order(meta$date, decreasing = T),  
  ]  
seqs <- meta[grep("P.1", meta$pango_lineage),  
  ]  
seqs_uy <- seqs[which(seqs$location ==  
  "Uruguay"), ]  
dim(seqs_uy) # UY seqs
```

```
## [1] 871    9
```

```
cl <- read.table("nextclade_out.tsv",  
  sep = "\t", header = T)  
cl <- cl[, 1:2]  
colnames(cl)[1] <- c("label")  
head(cl)
```

### Read nextclade output

```
##                                     label  
## 1          LV-012|EPI_ISL_437091|Europe|NA|B.1.1|2020-03-23  
## 2          EC-36230|EPI_ISL_491950|SouthAmerica|NA|B.1.1|2020-05-08  
## 3          EE-149467|EPI_ISL_1138530|Europe|NA|B.1.177.60|2021-01-26  
## 4          NL-ZH-RIVM-38556|EPI_ISL_2610019|Europe|NA|B.1.1.7|2021-05-27  
## 5 PE-CAL-INS-736|EPI_ISL_1138417|SouthAmerica|NA|B.1.1.348|2021-01-12  
## 6          EC-52438|EPI_ISL_491951|SouthAmerica|NA|B.1.1|2020-06-01  
##          clade  
## 1          20B  
## 2          20B  
## 3          20E (EU1)  
## 4 20I (Alpha, V1)  
## 5          20B  
## 6          20B
```

```
dim(cl)
```

```
## [1] 5176    2
```

```
t <- as.data.frame(as_tibble(tre))  
tt <- as.data.frame(str_split_fixed(t$label,  
  "[|]", 6))  
colnames(tt) <- c("id", "accession",  
  "region", "subregion", "lineage",
```

```

    "date")
t <- as.data.frame(cbind(t, tt))
head(t)

```

### Extract data from tree

```

##   parent node branch.length
## 1   5167    1      0.06411
## 2   5168    2      0.15978
## 3   5168    3      0.06411
## 4   5170    4      0.00546
## 5   5170    5      0.00000
## 6   5171    6      0.04331
##                                     label          id
## 1                CN-Hu-1|EPI_ISL_402125|Asia|NA|B|2019-12-26      CN-Hu-1
## 2            AU-VPRL014|EPI_ISL_455603|Oceania|NA|B.1|2020-01-30      AU-VPRL014
## 3                CN-WH01|EPI_ISL_406798|Asia|NA|B|2019-12-26      CN-WH01
## 4 AE-skmc-2200543130|EPI_ISL_582623|Asia|NA|B.53|2020-02-23 AE-skmc-2200543130
## 5 AE-skmc-2200522146|EPI_ISL_582611|Asia|NA|B.53|2020-02-21 AE-skmc-2200522146
## 6                SG-1|EPI_ISL_406973|Asia|NA|B|2020-01-23      SG-1
##   accession  region subregion lineage    date
## 1 EPI_ISL_402125   Asia      NA      B 2019-12-26
## 2 EPI_ISL_455603 Oceania      NA    B.1 2020-01-30
## 3 EPI_ISL_406798   Asia      NA      B 2019-12-26
## 4 EPI_ISL_582623   Asia      NA    B.53 2020-02-23
## 5 EPI_ISL_582611   Asia      NA    B.53 2020-02-21
## 6 EPI_ISL_406973   Asia      NA      B 2020-01-23

```

```
dim(t)
```

```
## [1] 9225  10
```

```
t <- left_join(t, cl, by = "label")
```

### Include clade data

```
pattern <- rle(t$region)
```

### Detect pattern continuity in the tree based on the region

```

tracks <- split(t$label, rep(seq_along(pattern$lengths),
  pattern$lengths))

```

### Extract tracks of continuity based on pattern

```
dl <- data.frame(ID = rep(names(tracks),
  apply(tracks, length)), Obs = unlist(tracks))

head(dl)
```

Transform list of vector into dataframe

```
##      ID                                                    Obs
## 1  1          CN-Hu-1|EPI_ISL_402125|Asia|NA|B|2019-12-26
## 2  2          AU-VPRL014|EPI_ISL_455603|Oceania|NA|B.1|2020-01-30
## 3  3          CN-WHO1|EPI_ISL_406798|Asia|NA|B|2019-12-26
## 4  3 AE-skmc-2200543130|EPI_ISL_582623|Asia|NA|B.53|2020-02-23
## 5  3 AE-skmc-2200522146|EPI_ISL_582611|Asia|NA|B.53|2020-02-21
## 6  3          SG-1|EPI_ISL_406973|Asia|NA|B|2020-01-23
```

```
uy <- dl[grepl("Uruguay", dl$Obs),
  ]
dim(uy)
```

Extract from the dataframe the Uruguayan sequences

```
## [1] 1792      2
```

```
head(uy)
```

```
##      ID                                                    Obs
## 23 11          UY-UY-10|EPI_ISL_429257|Uruguay|Montevideo|B.31|2020-03-19
## 28 16          UY-UY-NYUMC863|EPI_ISL_457959|Uruguay|Montevideo|B|2020-03-26
## 29 16          UY-UY-NYUMC865|EPI_ISL_457961|Uruguay|Montevideo|B|2020-03-28
## 35 21          UY-UY-7|EPI_ISL_426482|Uruguay|Montevideo|A.1|2020-03-17
## 36 21          UY-UY-8|EPI_ISL_426583|Uruguay|Montevideo|A.1|2020-03-17
## 39 24 UY-UY-NYUMC849|EPI_ISL_457945|Uruguay|Montevideo|A.2|2020-03-17
```

## Gamma introduction to Uruguay

Extract all tip labels based on pattern “P.1”

```
uy <- uy[grep("P.1", uy$Obs), ]
colnames(uy) <- c("cluster", "label")
```

Find the number of sequences per cluster

```
n <- uy %>%
  group_by(cluster) %>%
  summarise(count = n())
n <- n %>%
  filter(count > 1)
uy2 <- merge(n, uy, by = "cluster")
```

```
mrca <- uy2 %>%
  group_by(cluster) %>%
  summarise(count = n(), MRCA = MRCA(tre,
    label))
mrca <- as.data.frame(mrca)
mrca <- mrca[order(mrca$MRCA),
  ]
head(mrca)
```

Find the MRCA for each clustered sequences based on the tip labels in each cluster

```
##   cluster count MRCA
## 1    1264     2 8295
## 2    1268     2 8299
## 3    1270     2 8304
## 4    1274     2 8308
## 6    1290    392 8308
## 5    1280     2 8316
```

Find the tMRCA of the detected nodes in the dates.tsv output of the treetime joint analysis

```
target <- mrca$MRCA
nodes <- filter(t, node %in% target)
nodes <- select(nodes, parent,
  node, branch.length, label,
  id)
```

The node column corresponds to the MRCA

```
colnames(nodes)[2] <- c("MRCA")
mrca <- merge(mrca, nodes, by = "MRCA")
dat <- read.table("treetime_out_joint/dates.tsv",
  sep = "\t", header = F)
head(dat)
```

```
##                                     V1      V2      V3
```



```
## 1                                NODE_0000001 2019-12-03 2019.920824
## 2          CN-Hu-1|EPI_ISL_402125|Asia|NA|B|2019-12-26 2019-12-26 2019.984932
## 3                                NODE_0002994 2019-12-03 2019.920824
## 4 AU-VPRL014|EPI_ISL_455603|Oceania|NA|B.1|2020-01-30 2020-01-30 2020.080601
## 5          CN-WH01|EPI_ISL_406798|Asia|NA|B|2019-12-26 2019-12-26 2019.984932
## 6                                NODE_0000002 2019-12-22 2019.974710
```

```
tar <- mrca$label
n2 <- filter(dat, V1 %in% tar)
colnames(n2) <- c("label", "date",
  "numeric_date")
head(n2)
```

```
##          label          date numeric_date
## 1 NODE_0001140          --          --
## 2 NODE_0003892 2021-01-10 2021.025004
## 3 NODE_0003900 2021-02-23 2021.146464
## 4 NODE_0001142 2020-12-05 2020.928548
## 5 NODE_0001173 2021-07-27 2021.568928
## 6 NODE_0001413 2021-05-03 2021.334494
```

```
mrca <- merge(mrca, n2, by = "label")
mrca <- mrca[order(mrca$date, decreasing = F),
  ]
head(mrca)
```

```
##          label MRCA cluster count parent branch.length          id          date
## 1 NODE_0001140 8295      1264      2    8294      0.12544 NODE_0001140          --
## 2 NODE_0001142 8308      1274      2    8294      0.01547 NODE_0001142 2020-12-05
## 3 NODE_0001142 8308      1290     392    8294      0.01547 NODE_0001142 2020-12-05
## 21 NODE_0003892 8299      1268      2    8297      0.04501 NODE_0003892 2021-01-10
## 12 NODE_0001401 8651      1296      4    8650      0.05296 NODE_0001401 2021-01-26
## 18 NODE_0001826 8728      1330     55    8727      0.07232 NODE_0001826 2021-01-29
##          numeric_date
## 1          --
## 2    2020.928548
## 3    2020.928548
## 21    2021.025004
## 12    2021.070330
## 18    2021.079314
```

```
tre <- read.beast("treetime_out_joint/fixed_mugration_region/annotated_tree.nexus")
t <- as.data.frame(as_tibble(tre))
anc <- mrca$parent

loc <- filter(t, node %in% anc)
loc <- select(loc, node, location)
colnames(loc)[1] <- c("parent")

df <- left_join(mrca, loc, by = "parent")
```

Find the location annotation of the ancestor node of the MRCA of the UY cluster sequences

df

Gamma introduction tMRCA of UY clusters and number of introductions based on unique parent node

##	label	MRCA	cluster	count	parent	branch.length	id	date
## 1	NODE_0001140	8295	1264	2	8294	0.12544	NODE_0001140	--
## 2	NODE_0001142	8308	1274	2	8294	0.01547	NODE_0001142	2020-12-05
## 3	NODE_0001142	8308	1290	392	8294	0.01547	NODE_0001142	2020-12-05
## 4	NODE_0003892	8299	1268	2	8297	0.04501	NODE_0003892	2021-01-10
## 5	NODE_0001401	8651	1296	4	8650	0.05296	NODE_0001401	2021-01-26
## 6	NODE_0001826	8728	1330	55	8727	0.07232	NODE_0001826	2021-01-29
## 7	NODE_0004648	8827	1346	17	8817	0.01659	NODE_0004648	2021-01-29
## 8	NODE_0001431	8818	1344	12	8817	0.03020	NODE_0001431	2021-02-03
## 9	NODE_0004802	8962	1412	16	8958	0.00000	NODE_0004802	2021-02-10
## 10	NODE_0004812	8971	1426	222	8962	0.00000	NODE_0004812	2021-02-10
## 11	NODE_0001269	8927	1397	13	8926	0.03520	NODE_0001269	2021-02-12
## 12	NODE_0001906	8809	1339	3	8808	0.09057	NODE_0001906	2021-02-23
## 13	NODE_0003900	8304	1270	2	8302	0.00018	NODE_0003900	2021-02-23
## 14	NODE_0004817	8975	1414	27	8974	0.00000	NODE_0004817	2021-02-23
## 15	NODE_0004845	8999	1416	49	8987	0.01265	NODE_0004845	2021-02-27
## 16	NODE_0001143	8872	1360	4	8871	0.08424	NODE_0001143	2021-03-01
## 17	NODE_0004898	9043	1418	2	9042	0.00261	NODE_0004898	2021-03-09
## 18	NODE_0001214	8848	1351	7	8847	0.05102	NODE_0001214	2021-03-18
## 19	NODE_0001320	8913	1384	2	8912	0.09678	NODE_0001320	2021-03-27
## 20	NODE_0001447	8717	1322	2	8716	0.17191	NODE_0001447	2021-03-31
## 21	NODE_0001413	8645	1293	3	8643	0.12281	NODE_0001413	2021-05-03
## 22	NODE_0001818	8722	1325	2	8721	0.37430	NODE_0001818	2021-05-19
## 23	NODE_0001303	8924	1395	2	8923	0.25297	NODE_0001303	2021-06-03
## 24	NODE_0001887	8802	1336	2	8801	0.44453	NODE_0001887	2021-06-29
## 25	NODE_0001394	8878	1363	2	8877	0.30134	NODE_0001394	2021-06-30
## 26	NODE_0001173	8316	1280	2	8315	0.29904	NODE_0001173	2021-07-27
## 27	NODE_0001800	8692	1310	2	8691	0.30432	NODE_0001800	2021-07-30
## 28	NODE_0001365	8948	1408	3	8947	0.10338	NODE_0001365	2021-08-26
##	numeric_date	location						
## 1	--	Brazil						
## 2	2020.928548	Brazil						
## 3	2020.928548	Brazil						
## 4	2021.025004	Brazil						
## 5	2021.070330	Brazil						
## 6	2021.079314	Brazil						
## 7	2021.078167	Brazil						
## 8	2021.091779	Brazil						
## 9	2021.110799	Brazil						
## 10	2021.110799	Uruguay						
## 11	2021.116380	Brazil						
## 12	2021.146575	Brazil						
## 13	2021.146464	Brazil						
## 14	2021.146085	Uruguay						
## 15	2021.158735	Uruguay						
## 16	2021.164175	Brazil						
## 17	2021.185572	Uruguay						

```
## 18 2021.210171 Uruguay
## 19 2021.234338 Brazil
## 20 2021.244850 Brazil
## 21 2021.334494 Brazil
## 22 2021.380251 Brazil
## 23 2021.419489 Brazil
## 24 2021.491781 Brazil
## 25 2021.494521 Brazil
## 26 2021.568928 SouthAmerica
## 27 2021.576712 Argentina
## 28 2021.651746 Brazil
```

```
dim(as.data.frame(unique(df$parent)))
```

```
## [1] 25 1
```

```
write.table(df, "Gamma_intro.tsv",
  sep = "\t", row.names = F,
  quote = F)
```

```
# Function to subtract 1 from parent column if pattern "Uruguay" is found in location column
subtract_if_pattern_found <- function(df, pattern) {
  indices <- which(df$location == pattern)
  if (length(indices) > 0) {
    df$parent[indices] <- df$parent[indices] - 1
  }
  return(df)
}

# Load tree data
tre <-
  read.beast("treetime_out_joint/fixed_mugration_region/annotated_tree.nexus")
t <- as.data.frame(as_tibble(tre))

# Initialize counter
iterations <- 0

# Iterate until Uruguay pattern is no longer detected
repeat {
  # Apply function to modify data
  pattern <- "Uruguay"
  modified_data <- subtract_if_pattern_found(df, pattern)

  # Extract relevant data from tree
  anc <- modified_data$parent
  loc <- filter(t, node %in% anc)
  loc <- select(loc, node, location)
  colnames(loc)[1] <- c("parent")

  # Join modified data with location data
```

```

modified_data$location <- NULL
df <- left_join(modified_data, loc, by = "parent")

# Increment iteration counter
iterations <- iterations + 1

# Check if Uruguay pattern is no longer detected
if (!any(df$location == pattern)) {
  break
}
}

# Print final iteration count and modified data frame
print(paste("Number of iterations:", iterations))

```

Find the location of the ancestor outside Uruguay

```
## [1] "Number of iterations: 81"
```

```
print("Final Modified data:")
```

```
## [1] "Final Modified data:"
```

```
print(df)
```

##	label	MRCA	cluster	count	parent	branch.length	id	date
## 1	NODE_0001140	8295	1264	2	8294	0.12544	NODE_0001140	--
## 2	NODE_0001142	8308	1274	2	8294	0.01547	NODE_0001142	2020-12-05
## 3	NODE_0001142	8308	1290	392	8294	0.01547	NODE_0001142	2020-12-05
## 4	NODE_0003892	8299	1268	2	8297	0.04501	NODE_0003892	2021-01-10
## 5	NODE_0001401	8651	1296	4	8650	0.05296	NODE_0001401	2021-01-26
## 6	NODE_0001826	8728	1330	55	8727	0.07232	NODE_0001826	2021-01-29
## 7	NODE_0004648	8827	1346	17	8817	0.01659	NODE_0004648	2021-01-29
## 8	NODE_0001431	8818	1344	12	8817	0.03020	NODE_0001431	2021-02-03
## 9	NODE_0004802	8962	1412	16	8958	0.00000	NODE_0004802	2021-02-10
## 10	NODE_0004812	8971	1426	222	8961	0.00000	NODE_0004812	2021-02-10
## 11	NODE_0001269	8927	1397	13	8926	0.03520	NODE_0001269	2021-02-12
## 12	NODE_0001906	8809	1339	3	8808	0.09057	NODE_0001906	2021-02-23
## 13	NODE_0003900	8304	1270	2	8302	0.00018	NODE_0003900	2021-02-23
## 14	NODE_0004817	8975	1414	27	8961	0.00000	NODE_0004817	2021-02-23
## 15	NODE_0004845	8999	1416	49	8961	0.01265	NODE_0004845	2021-02-27
## 16	NODE_0001143	8872	1360	4	8871	0.08424	NODE_0001143	2021-03-01
## 17	NODE_0004898	9043	1418	2	8961	0.00261	NODE_0004898	2021-03-09
## 18	NODE_0001214	8848	1351	7	8846	0.05102	NODE_0001214	2021-03-18
## 19	NODE_0001320	8913	1384	2	8912	0.09678	NODE_0001320	2021-03-27
## 20	NODE_0001447	8717	1322	2	8716	0.17191	NODE_0001447	2021-03-31
## 21	NODE_0001413	8645	1293	3	8643	0.12281	NODE_0001413	2021-05-03
## 22	NODE_0001818	8722	1325	2	8721	0.37430	NODE_0001818	2021-05-19
## 23	NODE_0001303	8924	1395	2	8923	0.25297	NODE_0001303	2021-06-03
## 24	NODE_0001887	8802	1336	2	8801	0.44453	NODE_0001887	2021-06-29
## 25	NODE_0001394	8878	1363	2	8877	0.30134	NODE_0001394	2021-06-30
## 26	NODE_0001173	8316	1280	2	8315	0.29904	NODE_0001173	2021-07-27

```
## 27 NODE_0001800 8692 1310 2 8691 0.30432 NODE_0001800 2021-07-30
## 28 NODE_0001365 8948 1408 3 8947 0.10338 NODE_0001365 2021-08-26
## numeric_date location
## 1 -- Brazil
## 2 2020.928548 Brazil
## 3 2020.928548 Brazil
## 4 2021.025004 Brazil
## 5 2021.070330 Brazil
## 6 2021.079314 Brazil
## 7 2021.078167 Brazil
## 8 2021.091779 Brazil
## 9 2021.110799 Brazil
## 10 2021.110799 Argentina
## 11 2021.116380 Brazil
## 12 2021.146575 Brazil
## 13 2021.146464 Brazil
## 14 2021.146085 Argentina
## 15 2021.158735 Argentina
## 16 2021.164175 Brazil
## 17 2021.185572 Argentina
## 18 2021.210171 Brazil
## 19 2021.234338 Brazil
## 20 2021.244850 Brazil
## 21 2021.334494 Brazil
## 22 2021.380251 Brazil
## 23 2021.419489 Brazil
## 24 2021.491781 Brazil
## 25 2021.494521 Brazil
## 26 2021.568928 SouthAmerica
## 27 2021.576712 Argentina
## 28 2021.651746 Brazil
```

```
dim(as.data.frame(unique(df$parent)))
```

```
## [1] 22 1
```

```
table(df$location)
```

```
##
## Argentina Brazil SouthAmerica
## 5 22 1
```

```
phy <- tre@phylo
phy4 <- phylo4(phy)
desc <- phylobase::descendants(phy = phy4,
  node = 8294)
head(desc)
```

Find descendants of the largest clade

```
## PY-iics_12225|EPI_ISL_4071899|SouthAmerica|NA|P.1|2021-04-12
## 3991
## UY-CUY13-002914|EPI_ISL_NA|Uruguay|RioNegro|P.1|2021-06-06
## 3992
## UY-CUY13-002940|EPI_ISL_NA|Uruguay|RioNegro|P.1|2021-06-07
## 3993
## BR-PA-FIOCRUZ-14607|EPI_ISL_2645830|Brazil|NA|P.1|2021-03-09
## 3994
## BO-LaPaz_GM3-3|EPI_ISL_2802858|SouthAmerica|NA|P.1.14|2021-06-02
## 3995
## BR-CE-FIOCRUZ-00690|EPI_ISL_3912040|Brazil|NA|P.1|2021-01-20
## 3996
```

```
colnames(uy)[2] <- c("label")
colnames(meta)[1] <- c("label")
d <- merge(uy, meta, by = "label")
head(d)
```

```
##                                     label cluster
## 1          UY-CHY-M525|EPI_ISL_1991974|Uruguay|Rocha|P.1|2021-02-26    1344
## 2          UY-CHY-M531|EPI_ISL_1992100|Uruguay|Rocha|P.1|2021-02-23    1339
## 3 UY-CUY1-000011_UYAR|EPI_ISL_2031707|Uruguay|Artigas|P.1.10|2021-03-04    1369
## 4 UY-CUY1-000020_UYMO|EPI_ISL_2031708|Uruguay|Montevideo|P.1|2021-03-04    1426
## 5 UY-CUY1-000025_UYMO|EPI_ISL_2031709|Uruguay|Montevideo|P.1|2021-03-05    1426
## 6   UY-CUY1-000031_UYAR|EPI_ISL_2031710|Uruguay|Artigas|P.1|2021-03-05    1426
##      date      accession location      region country      city
## 1 2021-02-26 EPI_ISL_1991974  Uruguay South America Uruguay    Rocha
## 2 2021-02-23 EPI_ISL_1992100  Uruguay South America Uruguay    Rocha
## 3 2021-03-04 EPI_ISL_2031707  Uruguay South America Uruguay  Artigas
## 4 2021-03-04 EPI_ISL_2031708  Uruguay South America Uruguay Montevideo
## 5 2021-03-05 EPI_ISL_2031709  Uruguay South America Uruguay Montevideo
## 6 2021-03-05 EPI_ISL_2031710  Uruguay South America Uruguay  Artigas
##      subregion pango_lineage
## 1      Rocha          P.1
## 2      Rocha          P.1
## 3    Artigas      P.1.10
## 4 Montevideo          P.1
## 5 Montevideo          P.1
## 6    Artigas          P.1
```

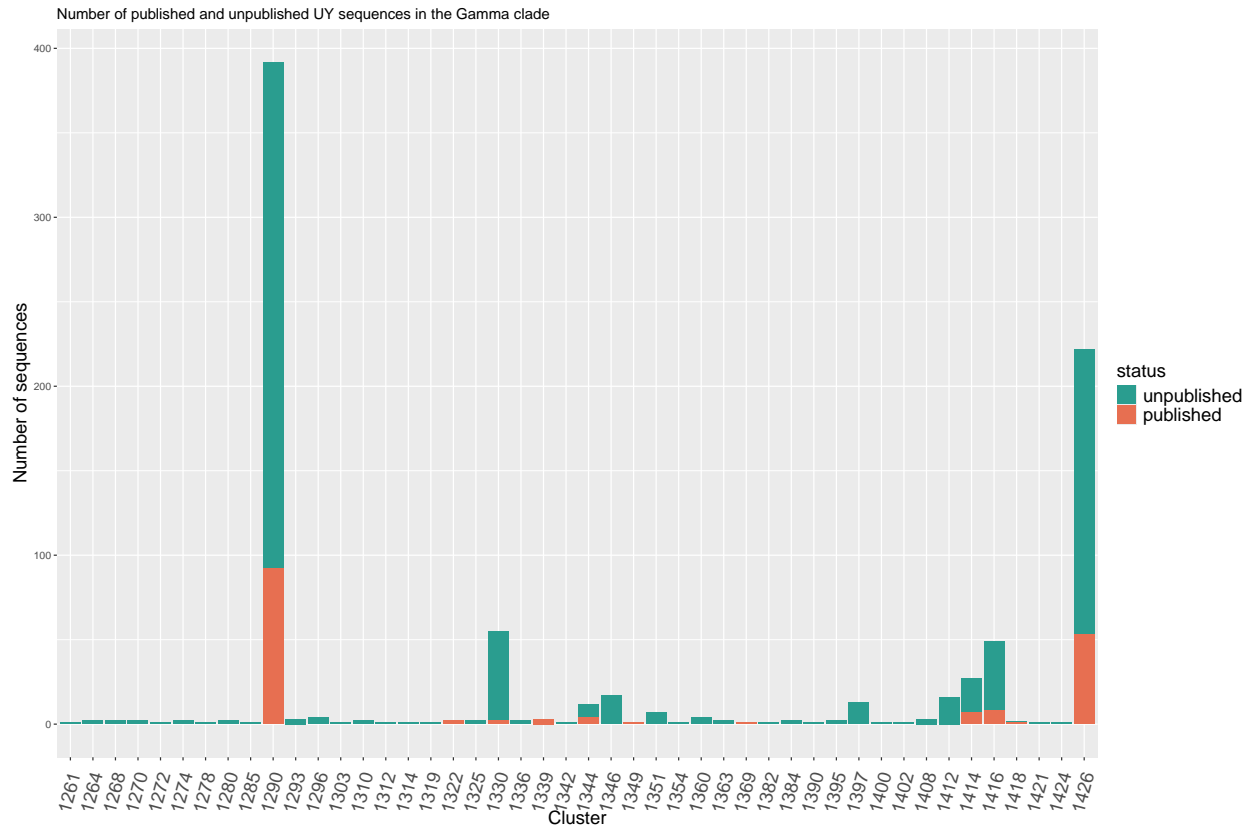
```
co2 <- d %>%
  mutate(status = ifelse(stringr::str_detect(accession,
    "EPI_ISL_NA"), "not_published",
    "published"))
st <- co2 %>%
  group_by(cluster, status) %>%
  summarise(count = n())
```

```
## 'summarise()' has grouped output by 'cluster'. You can override using the
## '.groups' argument.
```

```
st
```

```
## # A tibble: 53 x 3
## # Groups:   cluster [46]
##   cluster status      count
##   <chr>   <chr>      <int>
## 1 1261    not_published    1
## 2 1264    not_published    2
## 3 1268    not_published    2
## 4 1270    not_published    2
## 5 1272    not_published    1
## 6 1274    not_published    2
## 7 1278    not_published    1
## 8 1280    not_published    2
## 9 1285    not_published    1
## 10 1290   not_published   300
## # i 43 more rows
```

```
fig <- ggplot(st, aes(x = as.character(cluster),
  y = count, group = status,
  fill = status)) + geom_bar(stat = "identity") +
  scale_fill_manual(values = c(not_published = "#2a9d8f",
    published = "#e76f51"),
    labels = c("unpublished",
      "published")) + ylab("Number of sequences") +
  xlab("Cluster") + theme(axis.text.x = element_text(size = 16,
    angle = 75, vjust = 0.8, hjust = 1),
    legend.title = element_text(size = 16),
    legend.text = element_text(size = 16),
    axis.title = element_text(size = 16)) +
  ggtitle("Number of published and unpublished UY sequences in the Gamma clade")
fig
```



## Extract Gamma clade subtree

```
tre <- read.beast("treetime_out_joint/joint_mugration_region/annotated_tree.nexus")
t <- as.data.frame(as_tibble(tre))
head(t)
```

```
##   parent node branch.length
## 1    5167     1      0.06411
## 2    5168     2      0.15978
## 3    5168     3      0.06411
## 4    5170     4      0.00546
## 5    5170     5      0.00000
## 6    5171     6      0.04331
##
##                                     label location
## 1                CN-Hu-1|EPI_ISL_402125|Asia|NA|B|2019-12-26      Asia
## 2            AU-VPRL014|EPI_ISL_455603|Oceania|NA|B.1|2020-01-30  Oceania
## 3                CN-WH01|EPI_ISL_406798|Asia|NA|B|2019-12-26      Asia
## 4 AE-skmc-2200543130|EPI_ISL_582623|Asia|NA|B.53|2020-02-23      Asia
## 5 AE-skmc-2200522146|EPI_ISL_582611|Asia|NA|B.53|2020-02-21      Asia
## 6                SG-1|EPI_ISL_406973|Asia|NA|B|2020-01-23      Asia
```

```
tt <- as.data.frame(str_split_fixed(t$label,
  "[|]", 6))
colnames(tt) <- c("id", "accession",
```



```

      "region", "subregion", "lineage",
      "date")
t <- as.data.frame(cbind(t, tt))
head(t)

```

```

##      parent node branch.length
## 1      5167      1      0.06411
## 2      5168      2      0.15978
## 3      5168      3      0.06411
## 4      5170      4      0.00546
## 5      5170      5      0.00000
## 6      5171      6      0.04331
##
##                                     label location
## 1              CN-Hu-1|EPI_ISL_402125|Asia|NA|B|2019-12-26      Asia
## 2          AU-VPRL014|EPI_ISL_455603|Oceania|NA|B.1|2020-01-30  Oceania
## 3              CN-WH01|EPI_ISL_406798|Asia|NA|B|2019-12-26      Asia
## 4 AE-skmc-2200543130|EPI_ISL_582623|Asia|NA|B.53|2020-02-23      Asia
## 5 AE-skmc-2200522146|EPI_ISL_582611|Asia|NA|B.53|2020-02-21      Asia
## 6              SG-1|EPI_ISL_406973|Asia|NA|B|2020-01-23      Asia
##
##      id      accession      region subregion lineage      date
## 1      CN-Hu-1 EPI_ISL_402125      Asia      NA      B 2019-12-26
## 2      AU-VPRL014 EPI_ISL_455603  Oceania      NA      B.1 2020-01-30
## 3      CN-WH01 EPI_ISL_406798      Asia      NA      B 2019-12-26
## 4 AE-skmc-2200543130 EPI_ISL_582623      Asia      NA      B.53 2020-02-23
## 5 AE-skmc-2200522146 EPI_ISL_582611      Asia      NA      B.53 2020-02-21
## 6              SG-1 EPI_ISL_406973      Asia      NA      B 2020-01-23

```

```

cl <- read.table("nextclade_out.tsv",
  sep = "\t", header = T)
cl <- cl[, 1:2]
colnames(cl)[1] <- c("label")
head(cl)

```

```

##
##                                     label
## 1              LV-012|EPI_ISL_437091|Europe|NA|B.1.1|2020-03-23
## 2          EC-36230|EPI_ISL_491950|SouthAmerica|NA|B.1.1|2020-05-08
## 3          EE-149467|EPI_ISL_1138530|Europe|NA|B.1.177.60|2021-01-26
## 4      NL-ZH-RIVM-38556|EPI_ISL_2610019|Europe|NA|B.1.1.7|2021-05-27
## 5 PE-CAL-INS-736|EPI_ISL_1138417|SouthAmerica|NA|B.1.1.348|2021-01-12
## 6          EC-52438|EPI_ISL_491951|SouthAmerica|NA|B.1.1|2020-06-01
##
##      clade
## 1      20B
## 2      20B
## 3      20E (EU1)
## 4 20I (Alpha, V1)
## 5      20B
## 6      20B

```

```
dim(cl)
```

```
## [1] 5176      2
```

```
t <- merge(t, cl, by = "label")
```

```
unique(t$clade)
```

```
## [1] "21I (Delta)"      "21J (Delta)"      "20A"              "20H (Beta, V2)"
## [5] "20I (Alpha, V1)"  "19A"              "20C"              "20B"
## [9] "21A (Delta)"      "20D"              "21G (Lambda)"     "20J (Gamma, V3)"
## [13] "21H (Mu)"         "21K (Omicron)"    "21F (Iota)"       "21C (Epsilon)"
## [17] "19B"              "20E (EU1)"        "20F"              "21D (Eta)"
## [21] "recombinant"      "20G"
```

```
phy <- tre@phylo
```

```
gamma <- t[which(t$clade == "20J (Gamma, V3)",
  ]
gamma_l <- gamma$label
gamma_c <- findMRCA(phy, gamma_l)
gammatree <- extract.clade(phy,
  gamma_c)
t <- as_tibble(gammatree)
tt <- as.data.frame(str_split_fixed(t$label,
  "[|]", 6))
colnames(tt) <- c("id", "accession",
  "region", "subregion", "lineage",
  "date")
t <- as.data.frame(cbind(t, tt))
dim(t)
```

```
## [1] 2116 10
```

```
seqs_uy <- t[which(t$region ==
  "Uruguay"), ]
dim(seqs_uy) # UY seqs
```

```
## [1] 871 10
```

```
table(t$lineage)
```

```
##
##      P.1 P.1.1 P.1.10 P.1.12 P.1.14 P.1.15 P.1.2 P.1.4 P.1.6 P.1.7
##  936   826   215     5    18     5    64    19     3     7    16
## P.1.9
##      2
```

```
dir.create("gamma_subtree")
```

```
## Warning in dir.create("gamma_subtree"): 'gamma_subtree' ya existe
```

```
write.tree(gammatree, "gamma_subtree/gamma_subtree.nwk")
```

TreeTime was used to reconstruct ancestral states and to estimate the transition between regions

```
tre <- read.beast("gamma_subtree/joint_mugration_region_gamma/annotated_tree.nexus")
```

```
cl <- read.table("nextclade_out.tsv", sep = "\t",
  header = T)
cl <- cl[, 1:2]
colnames(cl)[1] <- c("label")
head(cl)
```

```
##                                     label
## 1          LV-012|EPI_ISL_437091|Europe|NA|B.1.1|2020-03-23
## 2          EC-36230|EPI_ISL_491950|SouthAmerica|NA|B.1.1|2020-05-08
## 3          EE-149467|EPI_ISL_1138530|Europe|NA|B.1.177.60|2021-01-26
## 4          NL-ZH-RIVM-38556|EPI_ISL_2610019|Europe|NA|B.1.1.7|2021-05-27
## 5 PE-CAL-INS-736|EPI_ISL_1138417|SouthAmerica|NA|B.1.1.348|2021-01-12
## 6          EC-52438|EPI_ISL_491951|SouthAmerica|NA|B.1.1|2020-06-01
##          clade
## 1          20B
## 2          20B
## 3          20E (EU1)
## 4 20I (Alpha, V1)
## 5          20B
## 6          20B
```

```
dim(cl)
```

```
## [1] 5176    2
```

```
t <- as.data.frame(as_tibble(tre))
tt <- as.data.frame(str_split_fixed(t$label,
  "[|]", 6))
colnames(tt) <- c("id", "accession", "region",
  "subregion", "lineage", "date")
t <- as.data.frame(cbind(t, tt))
head(t)
```

```
##   parent node branch.length
## 1   1182    1      0.61373
## 2   1184    2      0.37818
## 3   1185    3      0.22560
## 4   1185    4      0.12697
## 5   1186    5      0.89562
## 6   1187    6      0.36500
##                                     label
## 1 CO-ANT-CWOHC-VG-SEC00326N|EPI_ISL_2321578|SouthAmerica|NA|P.1|2021-03-26
## 2          UY-CUY19-004104|EPI_ISL_NA|Uruguay|Artigas|P.1.1|2021-07-21
```

```
## 3          BR-PB-FIOCRUZ-38030|EPI_ISL_3434905|Brazil|NA|P.1|2021-06-08
## 4          BR-PR-FIOCRUZ-28121|EPI_ISL_2557359|Brazil|NA|P.1|2021-05-03
## 5          PY-258289|EPI_ISL_6971786|SouthAmerica|NA|P.1|2021-10-07
## 6          PY-iics_12225|EPI_ISL_4071899|SouthAmerica|NA|P.1|2021-04-12
##      location              id      accession      region subregion
## 1 SouthAmerica CO-ANT-CWOHC-VG-SEC00326N EPI_ISL_2321578 SouthAmerica      NA
## 2      Uruguay      UY-CUY19-004104      EPI_ISL_NA      Uruguay      Artigas
## 3      Brazil      BR-PB-FIOCRUZ-38030 EPI_ISL_3434905      Brazil      NA
## 4      Brazil      BR-PR-FIOCRUZ-28121 EPI_ISL_2557359      Brazil      NA
## 5 SouthAmerica      PY-258289 EPI_ISL_6971786 SouthAmerica      NA
## 6 SouthAmerica      PY-iics_12225 EPI_ISL_4071899 SouthAmerica      NA
##      lineage      date
## 1      P.1 2021-03-26
## 2      P.1.1 2021-07-21
## 3      P.1 2021-06-08
## 4      P.1 2021-05-03
## 5      P.1 2021-10-07
## 6      P.1 2021-04-12
```

```
dim(t)
```

```
## [1] 2118    11
```

```
t <- merge(t, cl, by = "label")
```

```
uy <- t[which(t$region == "Uruguay"), ]
head(uy)
```

```
##                                     label
## 295          UY-CHY-M525|EPI_ISL_1991974|Uruguay|Rocha|P.1|2021-02-26
## 296          UY-CHY-M531|EPI_ISL_1992100|Uruguay|Rocha|P.1|2021-02-23
## 297 UY-CUY1-000011_UYAR|EPI_ISL_2031707|Uruguay|Artigas|P.1.10|2021-03-04
## 298 UY-CUY1-000020_UYMO|EPI_ISL_2031708|Uruguay|Montevideo|P.1|2021-03-04
## 299 UY-CUY1-000025_UYMO|EPI_ISL_2031709|Uruguay|Montevideo|P.1|2021-03-05
## 300      UY-CUY1-000031_UYAR|EPI_ISL_2031710|Uruguay|Artigas|P.1|2021-03-05
##      parent node branch.length location              id      accession
## 295      1713   655           0.03408 Uruguay      UY-CHY-M525 EPI_ISL_1991974
## 296      1702   645           0.00000 Uruguay      UY-CHY-M531 EPI_ISL_1992100
## 297      1778   751           0.04116 Uruguay UY-CUY1-000011_UYAR EPI_ISL_2031707
## 298      1973   996           0.04642 Uruguay UY-CUY1-000020_UYMO EPI_ISL_2031708
## 299      1990  1019           0.00000 Uruguay UY-CUY1-000025_UYMO EPI_ISL_2031709
## 300      2055  1112           0.01664 Uruguay UY-CUY1-000031_UYAR EPI_ISL_2031710
##      region  subregion lineage      date      clade
## 295 Uruguay      Rocha      P.1 2021-02-26 20J (Gamma, V3)
## 296 Uruguay      Rocha      P.1 2021-02-23 20J (Gamma, V3)
## 297 Uruguay      Artigas P.1.10 2021-03-04 20J (Gamma, V3)
## 298 Uruguay Montevideo      P.1 2021-03-04 20J (Gamma, V3)
## 299 Uruguay Montevideo      P.1 2021-03-05 20J (Gamma, V3)
## 300 Uruguay      Artigas      P.1 2021-03-05 20J (Gamma, V3)
```

```
dim(uy)
```

```
## [1] 871    12
```

```

uy <- uy[order(uy$date, decreasing = T),
]

new <- uy[which(uy$accession == "EPI_ISL_NA"),
]
dim(new)

```

```
## [1] 697 12
```

```

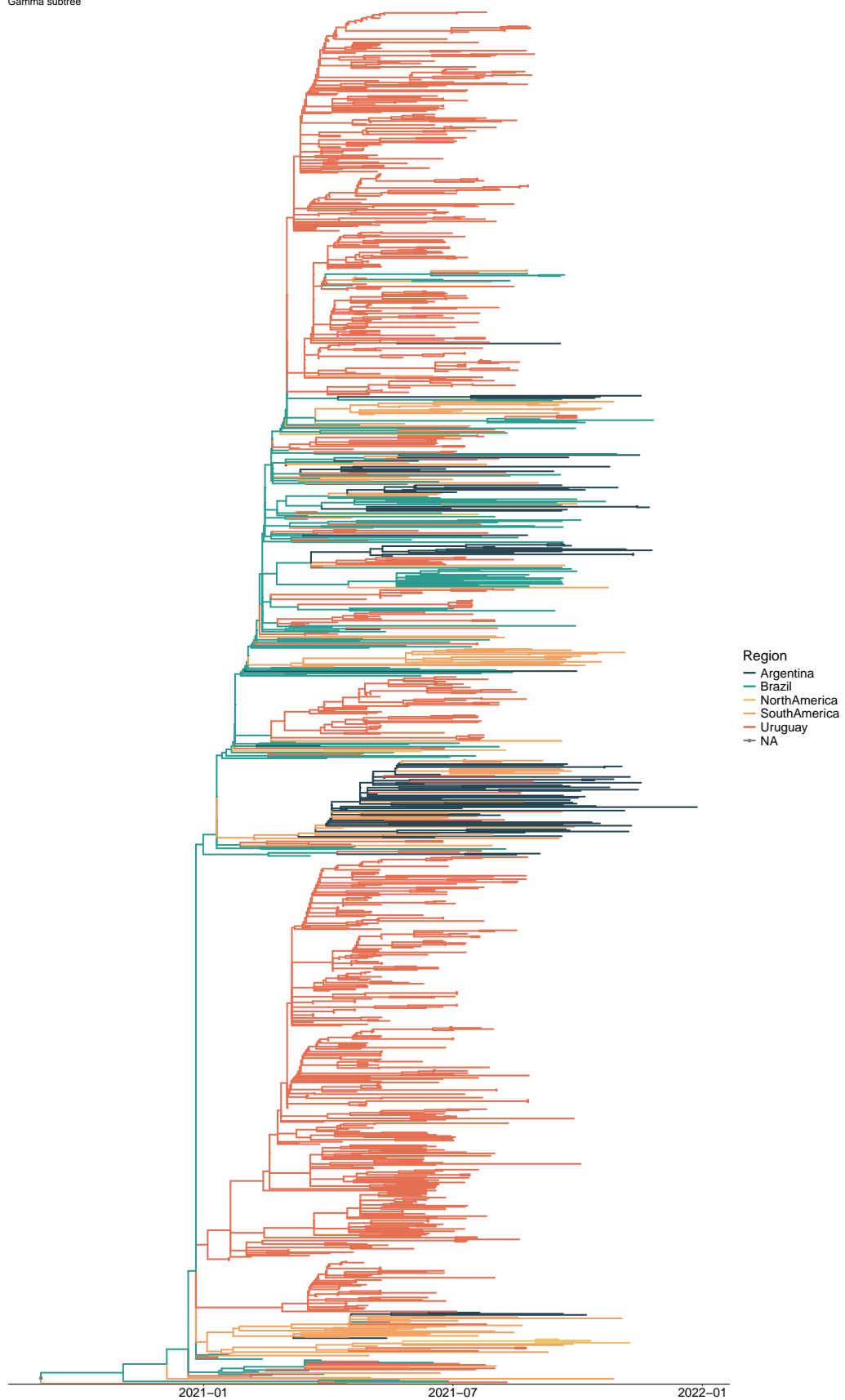
meta <- t

cls <- c("#264653", "#2a9d8f", "#e9c46a",
        "#f4a261", "#e76f51")

g <- ggtree(tre, aes(color = location), layout = "rectangular",
  size = 1, mrsd = "2021-12-28") + scale_color_manual(values = cls) +
  theme_tree2() + theme(axis.text = element_blank(),
    axis.line.y = element_blank()) + geom_rootpoint(position = "identity") +
  guides(color = guide_legend(title = "Region")) +
  scale_x_ggtree(labels = c("2020-01",
    "2020-07", "2021-01", "2021-07",
    "2022-01"), breaks = c(2020, 2020.5,
    2021, 2021.5, 2022)) + theme(axis.text = element_text(size = 16),
    legend.text = element_text(size = 16),
    legend.title = element_text(size = 18)) +
  ggtitle("Gamma subtree")
g

```

Gamma subtree



## Conclusions

The tMRCA of the Gamma sequences clusters from Uruguay dates from early December 2020. Based on unique parent nodes of the UY cluster, 25 introductions and re-introductions (?) were detected. The main source of introductions are located in Brazil. Two major clusters were detected, 396 sequences (node 8294, 2020-12-05, clusters 1290, 1274 y 1264) and 222 sequences (node 8961, 2021-02-10, cluster 1426), both clusters have a high proportion of new UY sequences.

## Create phylogeny

```
tre <- read.beast("treetime_out_joint/fixed_mugration_region/annotated_tree.nexus")

cl <- read.table("nextclade_out.tsv", sep = "\t",
  header = T)
cl <- cl[, 1:2]
colnames(cl)[1] <- c("label")
head(cl)
```

```
##                                     label
## 1          LV-012|EPI_ISL_437091|Europe|NA|B.1.1|2020-03-23
## 2          EC-36230|EPI_ISL_491950|SouthAmerica|NA|B.1.1|2020-05-08
## 3          EE-149467|EPI_ISL_1138530|Europe|NA|B.1.177.60|2021-01-26
## 4          NL-ZH-RIVM-38556|EPI_ISL_2610019|Europe|NA|B.1.1.7|2021-05-27
## 5 PE-CAL-INS-736|EPI_ISL_1138417|SouthAmerica|NA|B.1.1.348|2021-01-12
## 6          EC-52438|EPI_ISL_491951|SouthAmerica|NA|B.1.1|2020-06-01
##          clade
## 1          20B
## 2          20B
## 3          20E (EU1)
## 4 20I (Alpha, V1)
## 5          20B
## 6          20B
```

```
dim(cl)
```

```
## [1] 5176    2
```

```
t <- as.data.frame(as_tibble(tre))
tt <- as.data.frame(str_split_fixed(t$label,
  "[|]", 6))
colnames(tt) <- c("id", "accession", "region",
  "subregion", "lineage", "date")
t <- as.data.frame(cbind(t, tt))
head(t)
```

```
##   parent node branch.length
## 1    5167    1         0.06411
```

```
## 2 5168 2 0.15978
## 3 5168 3 0.06411
## 4 5170 4 0.00546
## 5 5170 5 0.00000
## 6 5171 6 0.04331
##
##                                label location
## 1 CN-Hu-1|EPI_ISL_402125|Asia|NA|B|2019-12-26 Asia
## 2 AU-VPRL014|EPI_ISL_455603|Oceania|NA|B.1|2020-01-30 Oceania
## 3 CN-WHO1|EPI_ISL_406798|Asia|NA|B|2019-12-26 Asia
## 4 AE-skmc-2200543130|EPI_ISL_582623|Asia|NA|B.53|2020-02-23 Asia
## 5 AE-skmc-2200522146|EPI_ISL_582611|Asia|NA|B.53|2020-02-21 Asia
## 6 SG-1|EPI_ISL_406973|Asia|NA|B|2020-01-23 Asia
##
##      id      accession region subregion lineage      date
## 1 CN-Hu-1 EPI_ISL_402125 Asia NA B 2019-12-26
## 2 AU-VPRL014 EPI_ISL_455603 Oceania NA B.1 2020-01-30
## 3 CN-WHO1 EPI_ISL_406798 Asia NA B 2019-12-26
## 4 AE-skmc-2200543130 EPI_ISL_582623 Asia NA B.53 2020-02-23
## 5 AE-skmc-2200522146 EPI_ISL_582611 Asia NA B.53 2020-02-21
## 6 SG-1 EPI_ISL_406973 Asia NA B 2020-01-23
```

```
dim(t)
```

```
## [1] 9225 11
```

```
t <- merge(t, cl, by = "label")
t <- t[order(t$date, decreasing = T), ]
head(t)
```

```
##
##                                label parent
## 1208 CN-4044|EPI_ISL_8187354|Asia|NA|BA.1|2021-12-28 7713
## 1939 GF-IPG202101560|EPI_ISL_8207166|SouthAmerica|NA|AY.99.2|2021-12-25 6075
## 1940 GF-IPG202101562|EPI_ISL_8207168|SouthAmerica|NA|AY.43|2021-12-25 6773
## 2025 HK-VM21048074|EPI_ISL_8189300|Asia|NA|BA.1|2021-12-25 7685
## 2320 MA-165|EPI_ISL_8144256|Africa|NA|AY.98|2021-12-25 6307
## 567 BR-CE-FIOCRUZ-45034CE|EPI_ISL_8184739|Brazil|NA|BA.1|2021-12-24 7686
##
##      node branch.length      location      id      accession
## 1208 3309      0.09819 Asia CN-4044 EPI_ISL_8187354
## 1939 1225      0.26301 SouthAmerica GF-IPG202101560 EPI_ISL_8207166
## 1940 2132      0.34359 SouthAmerica GF-IPG202101562 EPI_ISL_8207168
## 2025 3274      0.08493 Asia HK-VM21048074 EPI_ISL_8189300
## 2320 1538      0.65490 Africa MA-165 EPI_ISL_8144256
## 567 3279      0.04932 Brazil BR-CE-FIOCRUZ-45034CE EPI_ISL_8184739
##
##      region subregion lineage      date      clade
## 1208 Asia NA BA.1 2021-12-28 21K (Omicron)
## 1939 SouthAmerica NA AY.99.2 2021-12-25 21J (Delta)
## 1940 SouthAmerica NA AY.43 2021-12-25 21J (Delta)
## 2025 Asia NA BA.1 2021-12-25 21K (Omicron)
## 2320 Africa NA AY.98 2021-12-25 21J (Delta)
## 567 Brazil NA BA.1 2021-12-24 21K (Omicron)
```

```
unique(t$clade)
```



```
## [1] "21K (Omicron)" "21J (Delta)" "21A (Delta)" "21I (Delta)"
## [5] "21H (Mu)" "20J (Gamma, V3)" "21G (Lambda)" "20A"
## [9] "20I (Alpha, V1)" "20C" "20B" "21D (Eta)"
## [13] "21F (Iota)" "19B" "20H (Beta, V2)" "20D"
## [17] "recombinant" "21C (Epsilon)" "20G" "20E (EU1)"
## [21] "19A" "20F"
```

```
uy <- t[which(t$region == "Uruguay"), ]
dim(uy)
```

```
## [1] 1792 12
```

```
meta <- t

cls <- c("#f94144", "#f3722c", "#f8961e",
        "#f9844a", "#f9c74f", "#90be6d", "#43aa8b",
        "#4d908e", "#577590", "#277da1")

g <- ggtree(tre, aes(color = location), layout = "rectangular",
  size = 1, mrsd = "2021-12-28") + scale_color_manual(values = cls) +
  theme_tree2() + theme(axis.text = element_blank(),
    axis.line.y = element_blank()) + geom_rootpoint(position = "identity") +
  guides(color = guide_legend(title = "Region")) +
  scale_x_ggtree(labels = c("2020-01",
    "2020-03", "2020-07", "2020-10",
    "2021-01", "2021-03", "2021-07",
    "2021-10", "2022-01"), breaks = c(2020,
    2020.25, 2020.5, 2020.75, 2021, 2021.25,
    2021.5, 2021.75, 2022)) + theme(axis.text = element_text(size = 16),
    legend.text = element_text(size = 16),
    legend.title = element_text(size = 18))
```

```
p1 <- t[which(t$clade == "20J (Gamma, V3)"),
  ]
p1_lab <- p1$label

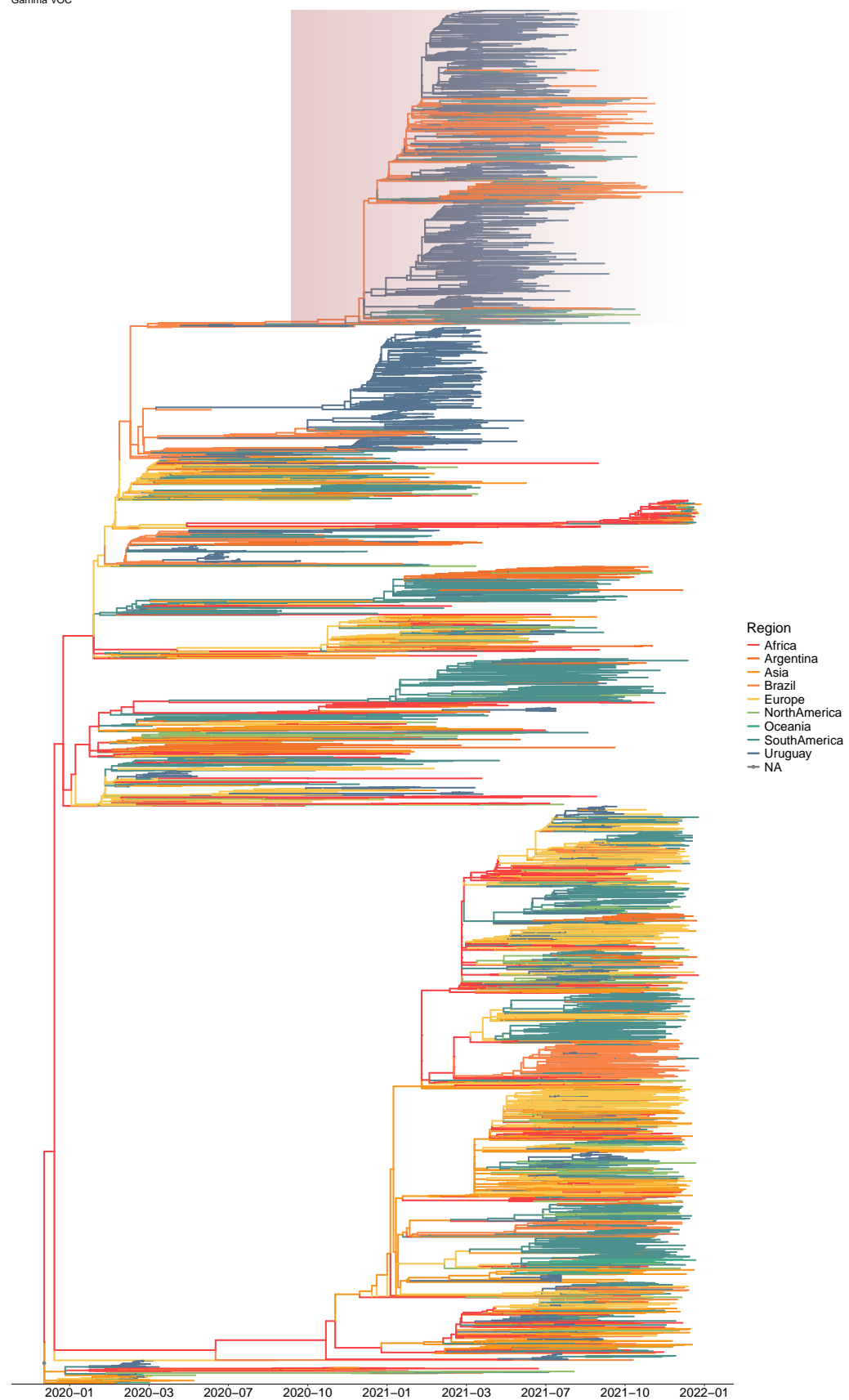
phy <- tre@phylo
gamma_c <- findMRCA(phy, gamma_l)
gamma_f <- c("Clade")
dt <- as.data.frame(gamma_c, gamma_f)

g1 <- g + new_scale_fill()

cls <- c("#ae4d51")

fig <- g1 + geom_hilight(data = dt,
  aes(node = gamma_c), gradient = T,
  alpha = 0.3, fill = cls) +
  ggtitle("Gamma VOC")
fig
```

Gamma VOC



## sessionInfo()

```
## R version 4.3.3 (2024-02-29)
## Platform: x86_64-redhat-linux-gnu (64-bit)
## Running under: Fedora Linux 39 (Workstation Edition)
##
## Matrix products: default
## BLAS/LAPACK: FlexiBLAS OPENBLAS-OPENMP; LAPACK version 3.11.0
##
## locale:
##  [1] LC_CTYPE=es_UY.UTF-8      LC_NUMERIC=C
##  [3] LC_TIME=es_UY.UTF-8      LC_COLLATE=es_UY.UTF-8
##  [5] LC_MONETARY=es_UY.UTF-8  LC_MESSAGES=es_UY.UTF-8
##  [7] LC_PAPER=es_UY.UTF-8     LC_NAME=C
##  [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=es_UY.UTF-8 LC_IDENTIFICATION=C
##
## time zone: America/Montevideo
## tzcode source: system (glibc)
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
##  [1] phylobase_0.8.12  phytools_2.2-1    maps_3.4.2        RColorBrewer_1.1-3
##  [5] dplyr_1.1.4       stringr_1.5.1     ggnewscale_0.4.10 tidytree_0.4.6
##  [9] phylotate_1.3     ggplot2_3.5.0     ggtree_3.10.1     treeio_1.26.0
## [13] ape_5.7-1
##
## loaded via a namespace (and not attached):
##  [1] ade4_1.7-22        tidyselect_1.2.1    farver_2.1.1
##  [4] optimParallel_1.0-2 fastmap_1.1.1        lazyeval_0.2.2
##  [7] combinat_0.0-8     XML_3.99-0.16.1     digest_0.6.35
## [10] lifecycle_1.0.4    magrittr_2.0.3      compiler_4.3.3
## [13] progress_1.2.3     rlang_1.1.3         tools_4.3.3
## [16] igraph_2.0.3       utf8_1.2.4          yaml_2.3.8
## [19] knitr_1.45         phangorn_2.11.1     clusterGeneration_1.3.8
## [22] labeling_0.4.3     prettyunits_1.2.0   mnormt_2.1.1
## [25] scatterplot3d_0.3-44 xml2_1.3.6          plyr_1.8.9
## [28] RNeXML_2.4.11      aplot_0.2.2         expm_0.999-9
## [31] withr_3.0.0        purrr_1.0.2         numDeriv_2016.8-1.1
## [34] grid_4.3.3         fansi_1.0.6         colorspace_2.1-0
## [37] scales_1.3.0       iterators_1.0.14    MASS_7.3-60.0.1
## [40] cli_3.6.2          crayon_1.5.2        rmarkdown_2.26
## [43] generics_0.1.3     rstudioapi_0.15.0   httr_1.4.7
## [46] reshape2_1.4.4     cachem_1.0.8        parallel_4.3.3
## [49] ggplotify_0.1.2    formatR_1.14        yulab.utils_0.1.4
## [52] vctr_0.6.5         Matrix_1.6-5        jsonlite_1.8.8
## [55] hms_1.1.3          gridGraphics_0.5-1  patchwork_1.2.0
## [58] foreach_1.5.2      tidyr_1.3.1         glue_1.7.0
## [61] codetools_0.2-19   DEoptim_2.2-8       stringi_1.8.3
## [64] gtable_0.3.4       quadprog_1.5-8      munsell_0.5.0
## [67] tibble_3.2.1       pillar_1.9.0        htmltools_0.5.7
```

## [70] R6_2.5.1	doParallel_1.0.17	evaluate_0.23
## [73] lattice_0.22-5	highr_0.10	memoise_2.0.1
## [76] ggfun_0.1.4	rncl_0.8.7	Rcpp_1.0.12
## [79] uuid_1.2-0	fastmatch_1.1-4	coda_0.19-4.1
## [82] nlme_3.1-164	xfun_0.42	fs_1.6.3
## [85] pkgconfig_2.0.3		