Introduction of Delta VOC 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 reagion 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 retrived 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 laoded 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
##
## Guangchuang Yu. Using ggtree to visualize data on tree-like structures.
## Current Protocols in Bioinformatics. 2020, 69:e96. doi:10.1002/cpbi.96
##
## Guangchuang Yu.  Data Integration, Manipulation and Visualization of
## Phylogenetic Trees (1st edition). Chapman and Hall/CRC. 2022,
## doi:10.1201/9781003279242
```

```r
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
##
## 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
##
## 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
```

```
##
## Attaching package: 'ggtree'
```

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

```r
library(ggplot2)
library(ape)
library(treeio)
library(phylotate)
library(tidytree)
```

```
## If you use the ggtree package suite in published research, please cite
```

```
## the appropriate paper(s):
##
## Guangchuang Yu. Using ggtree to visualize data on tree-like structures.
## Current Protocols in Bioinformatics. 2020, 69:e96. doi:10.1002/cpbi.96
##
## Guangchuang Yu.  Data Integration, Manipulation and Visualization of
## Phylogenetic Trees (1st edition). Chapman and Hall/CRC. 2022,
## doi:10.1201/9781003279242


##
## 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
```

```r
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
```

```r
library(RColorBrewer)
library(phytools)
```

```
## Loading required package: maps

##
## Attaching package: 'phytools'
```

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

```r
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
```

```r
library(phylotate)
```

```r
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
```

```r
dim(meta)
```

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

```r
meta <- meta[order(meta$date, decreasing = T),
    ]
```

```r
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
```

```r
dim(cl)
```

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

```r
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
```

```r
dim(t)
```

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

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

Include clade data

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

Detect pattern continuity in the tree based on the region

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

Extract tracks of continuity based on pattern

```r
dl <- data.frame(ID = rep(names(tracks),
    sapply(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-WH01|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
```

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

Extract from the dataframe the Uruguayan sequences

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

```r
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
```

Delta introduction to Uruguay

```r
dll <- as.data.frame(str_split_fixed(dl$Obs, "[|]", 6))
colnames(dll) <- c("id", "accession", "region", "subregion", "lineage", "date")
dl <- as.data.frame(cbind(dl, dll))
head(dl)
```

Include clade information to the dataframe containing the clusters

```
##   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-WH01|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
##                    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
```

```r
colnames(cl)[1] <- c("Obs")
head(cl)
```

```
##                                                                  Obs
## 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
```

```r
dl <- merge(dl, cl, by = "Obs")
head(dl)
```

```
##                                                                      Obs
## 1   AD-AND-242_212581387701_COV-GC|EPI_ISL_5684463|Europe|NA|AY.9|2021-09-11
## 2 AD-AND-245_212871383801_COV-GC|EPI_ISL_6125409|Europe|NA|AY.124|2021-10-09
## 3 AD-AND-245_213501452101_COV-GC|EPI_ISL_8185538|Europe|NA|AY.125|2021-12-12
## 4 AD-AND-245_213501452301_COV-GC|EPI_ISL_8185535|Europe|NA|AY.122|2021-12-10
## 5  AD-AND-247_212871382801_COV-GC|EPI_ISL_6125408|Europe|NA|AY.42|2021-09-29
## 6 AD-AND-247_213001322901_COV-GC|EPI_ISL_7589517|Europe|NA|AY.124|2021-10-16
##    ID                             id       accession region subregion lineage
## 1 114 AD-AND-242_212581387701_COV-GC EPI_ISL_5684463 Europe        NA    AY.9
## 2 240 AD-AND-245_212871383801_COV-GC EPI_ISL_6125409 Europe        NA  AY.124
## 3 299 AD-AND-245_213501452101_COV-GC EPI_ISL_8185538 Europe        NA  AY.125
## 4 569 AD-AND-245_213501452301_COV-GC EPI_ISL_8185535 Europe        NA  AY.122
## 5 256 AD-AND-247_212871382801_COV-GC EPI_ISL_6125408 Europe        NA   AY.42
## 6 240 AD-AND-247_213001322901_COV-GC EPI_ISL_7589517 Europe        NA  AY.124
##         date       clade
## 1 2021-09-11 21I (Delta)
## 2 2021-10-09 21J (Delta)
```

```
## 3 2021-12-12 21J (Delta)
## 4 2021-12-10 21J (Delta)
## 5 2021-09-29 21J (Delta)
## 6 2021-10-16 21J (Delta)
```

```r
dim(dl)
```

```
## [1] 5166    9
```

```r
target <- c("21J (Delta)", "21A (Delta)", "21I (Delta)")

d <- filter(dl, clade %in% target)
dim(d)
```

Based on Nextrain clade, extract all Delta sequences from Uruguay

```
## [1] 2083    9
```

```r
unique(d$region)
```

```
## [1] "Europe"       "NorthAmerica" "Africa"       "Argentina"    "Oceania"
## [6] "SouthAmerica" "Asia"         "Brazil"       "Uruguay"
```

```r
d <- d[which(d$region == "Uruguay"),]
head(d)
```

```
##                                                                          Obs  ID
## 1773 UY-CUY16-003792|EPI_ISL_NA|Uruguay|Montevideo|B.1.617.2|2021-07-05   55
## 1774  UY-CUY17-003842|EPI_ISL_NA|Uruguay|Lavalleja|B.1.617.2|2021-07-05  446
## 1775 UY-CUY17-003872|EPI_ISL_NA|Uruguay|Montevideo|B.1.617.2|2021-07-14  553
## 1776          UY-CUY17-003887|EPI_ISL_NA|Uruguay|NA|B.1.617.2|2021-07-14  593
## 1777          UY-CUY17-003888|EPI_ISL_NA|Uruguay|NA|B.1.617.2|2021-07-14  593
## 1778          UY-CUY17-003889|EPI_ISL_NA|Uruguay|NA|B.1.617.2|2021-07-14  595
##                   id accession  region  subregion   lineage       date
## 1773 UY-CUY16-003792 EPI_ISL_NA Uruguay Montevideo B.1.617.2 2021-07-05
## 1774 UY-CUY17-003842 EPI_ISL_NA Uruguay  Lavalleja B.1.617.2 2021-07-05
## 1775 UY-CUY17-003872 EPI_ISL_NA Uruguay Montevideo B.1.617.2 2021-07-14
## 1776 UY-CUY17-003887 EPI_ISL_NA Uruguay         NA B.1.617.2 2021-07-14
## 1777 UY-CUY17-003888 EPI_ISL_NA Uruguay         NA B.1.617.2 2021-07-14
## 1778 UY-CUY17-003889 EPI_ISL_NA Uruguay         NA B.1.617.2 2021-07-14
##           clade
## 1773 21A (Delta)
## 1774 21J (Delta)
## 1775 21J (Delta)
## 1776 21J (Delta)
## 1777 21J (Delta)
## 1778 21J (Delta)
```

```r
dim(d)
```

```
## [1] 280   9
```

```r
unique(d$clade)
```

```
## [1] "21A (Delta)" "21J (Delta)" "21I (Delta)"
```

```r
uy <- select(d, ID, Obs)
colnames(uy) <- c("cluster", "label")
head(uy)
```

```
##      cluster                                                              label
## 1773      55 UY-CUY16-003792|EPI_ISL_NA|Uruguay|Montevideo|B.1.617.2|2021-07-05
## 1774     446  UY-CUY17-003842|EPI_ISL_NA|Uruguay|Lavalleja|B.1.617.2|2021-07-05
## 1775     553 UY-CUY17-003872|EPI_ISL_NA|Uruguay|Montevideo|B.1.617.2|2021-07-14
## 1776     593         UY-CUY17-003887|EPI_ISL_NA|Uruguay|NA|B.1.617.2|2021-07-14
## 1777     593         UY-CUY17-003888|EPI_ISL_NA|Uruguay|NA|B.1.617.2|2021-07-14
## 1778     595         UY-CUY17-003889|EPI_ISL_NA|Uruguay|NA|B.1.617.2|2021-07-14
```

```r
unique(uy$cluster)
```

```
##  [1] "55"  "446" "553" "593" "595" "95"  "40"  "381" "43"  "548" "502" "244"
## [13] "48"  "579" "490" "601" "265" "283" "210" "228" "296" "76"  "159" "157"
## [25] "71"  "689" "285" "97"  "115" "561" "521" "492" "469" "649" "703" "292"
## [37] "667" "707" "90"  "651" "655" "507" "709" "423" "589" "150" "572" "177"
## [49] "597" "117" "298" "101" "694" "86"  "498" "418" "591" "483" "203" "412"
## [61] "92"  "148" "294" "676" "583" "152" "711" "480" "154" "65"  "653" "555"
## [73] "716" "700" "684" "334" "438" "277" "372" "349" "713" "692" "208" "182"
```

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

```
##    cluster count MRCA
## 14      40     9 5235
## 17      43     8 5247
## 19      48     2 5259
## 36      65     5 5293
## 44      76     4 5308
## 45      86     2 5330
```

```r
sum(mrca$count)
```

```
## [1] 245
```

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

colnames(nodes)[2] <- c("MRCA")
mrca <- merge(mrca, nodes, by = "MRCA")
```

```r
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
```

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

```
##          label       date numeric_date
## 1 NODE_0003085 2021-06-17   2021.459229
## 2 NODE_0000081 2021-07-11   2021.524908
## 3 NODE_0003120 2021-07-15   2021.535616
## 4 NODE_0003157 2021-06-27   2021.486376
## 5 NODE_0003174 2021-05-24   2021.394269
## 6 NODE_0003244 2021-08-03   2021.588492
```

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

```
##            label MRCA cluster count parent branch.length       date numeric_date
## 15 NODE_0003292 5337      90     2   5332       0.05234 2021-05-01   2021.331449
## 36 NODE_0004568 6466     572     5   6465       0.02861 2021-05-14   2021.366488
## 16 NODE_0003294 5338      92     2   5337       0.03650 2021-05-15   2021.367948
## 17 NODE_0003297 5340      95     6   5338       0.02167 2021-05-23   2021.389620
## 12 NODE_0003174 5308      76     4   5297       0.17550 2021-05-24   2021.394269
## 37 NODE_0004601 6495     579     5   6465       0.09553 2021-06-08   2021.433405
```

```r
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")
```

```
df
```

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

```
##           label MRCA cluster count parent branch.length       date numeric_date
## 1  NODE_0003292 5337      90     2   5332       0.05234 2021-05-01  2021.331449
## 2  NODE_0004568 6466     572     5   6465       0.02861 2021-05-14  2021.366488
## 3  NODE_0003294 5338      92     2   5337       0.03650 2021-05-15  2021.367948
## 4  NODE_0003297 5340      95     6   5338       0.02167 2021-05-23  2021.389620
## 5  NODE_0003174 5308      76     4   5297       0.17550 2021-05-24  2021.394269
## 6  NODE_0004601 6495     579     5   6465       0.09553 2021-06-08  2021.433405
## 7  NODE_0003085 5235      40     9   5234       0.10765 2021-06-17  2021.459229
## 8  NODE_0003684 5760     265     5   5759       0.02676 2021-06-23  2021.474940
## 9  NODE_0003157 5293      65     5   5290       0.04103 2021-06-27  2021.486376
## 10 NODE_0003955 5454     159    18   5453       0.12267 2021-06-27  2021.486315
## 11 NODE_0003313 5351     101     2   5347       0.06543 2021-06-28  2021.490373
## 12 NODE_0004602 6539     593     6   6505       0.03939 2021-07-02  2021.499107
## 13 NODE_0004041 5632     210     3   5631       0.10060 2021-07-03  2021.503097
## 14 NODE_0004967 6429     553     2   6428       0.12285 2021-07-06  2021.511948
## 15 NODE_0003949 5448     157     6   5447       0.43855 2021-07-07  2021.513508
## 16 NODE_0005085 6571     601     2   6570       0.19990 2021-07-08  2021.516438
## 17 NODE_0003476 5780     283     5   5779       0.06914 2021-07-09  2021.519971
## 18 NODE_0004943 6419     548     3   6418       0.13712 2021-07-09  2021.519178
## 19 NODE_0000081 5247      43     8   5243       0.00236 2021-07-11  2021.524908
## 20 NODE_0001607 6550     595     2   6549       0.00235 2021-07-13  2021.530137
## 21 NODE_0003394 5969     381     4   5968       0.01505 2021-07-13  2021.530137
## 22 NODE_0003894 5424     150     4   5414       0.04240 2021-07-14  2021.533856
## 23 NODE_0003120 5259      48     2   5257       0.35181 2021-07-15  2021.535616
## 24 NODE_0004504 6356     521     3   6355       0.04150 2021-07-15  2021.535616
## 25 NODE_0005142 6281     469     4   6280       0.25109 2021-07-15  2021.535616
## 26 NODE_0005000 6437     561     3   6433       0.11709 2021-07-18  2021.545204
## 27 NODE_0003222 5380     115     2   5379       0.06480 2021-07-19  2021.546575
## 28 NODE_0003627 5827     296     2   5826       0.10458 2021-07-19  2021.546575
## 29 NODE_0003307 5348      97     2   5347       0.12711 2021-07-21  2021.552055
## 30 NODE_0003495 5794     292     2   5793       0.06171 2021-07-21  2021.552055
## 31 NODE_0004527 6333     507     5   6330       0.06827 2021-07-21  2021.552073
## 32 NODE_0003484 5787     285     4   5785       0.03175 2021-07-23  2021.557534
## 33 NODE_0004163 6085     418     4   6082       0.03805 2021-07-23  2021.558013
## 34 NODE_0001884 6775     711    13   6772       0.00243 2021-07-25  2021.563791
## 35 NODE_0001308 6294     480     5   6293       0.09123 2021-07-26  2021.565565
## 36 NODE_0001888 6777     707     2   6776       0.00784 2021-07-29  2021.573973
## 37 NODE_0004660 6519     589     5   6518       0.04286 2021-07-29  2021.573932
```

```
## 38 NODE_0004282 6121    423     2  6120    0.21243 2021-07-30  2021.576712
## 39 NODE_0001616 6555    597     2  6553    0.00245 2021-07-31  2021.579452
## 40 NODE_0001891 6779    709     4  6778    0.01096 2021-07-31  2021.579452
## 41 NODE_0004544 6317    498     9  6309    0.13377 2021-08-01  2021.582784
## 42 NODE_0003501 5797    294    30  5791    0.13816 2021-08-02  2021.585769
## 43 NODE_0003244 5330     86     2  5329    0.16947 2021-08-03  2021.588492
## 44 NODE_0001577 6525    591     2  6524    0.00249 2021-08-07  2021.598630
## 45 NODE_0004873 6738    694     7  6737    0.08791 2021-08-07  2021.598630
## 46 NODE_0005018 6299    483     2  6298    0.08400 2021-08-08  2021.601370
## 47 NODE_0004869 6764    700     4  6763    0.07476 2021-08-09  2021.602747
## 48 NODE_0004893 6798    716    13  6795    0.03275 2021-08-21  2021.636149
## 49 NODE_0004740 6687    684     2  6686    0.03823 2021-09-07  2021.682646
##        location
## 1        Africa
## 2  SouthAmerica
## 3       Uruguay
## 4       Uruguay
## 5        Africa
## 6  SouthAmerica
## 7        Africa
## 8        Europe
## 9          Asia
## 10         Asia
## 11      Uruguay
## 12 SouthAmerica
## 13 SouthAmerica
## 14       Europe
## 15         Asia
## 16       Europe
## 17 SouthAmerica
## 18       Europe
## 19      Uruguay
## 20 SouthAmerica
## 21       Europe
## 22       Europe
## 23       Africa
## 24 SouthAmerica
## 25       Africa
## 26       Europe
## 27       Europe
## 28       Europe
## 29      Uruguay
## 30 SouthAmerica
## 31      Uruguay
## 32 SouthAmerica
## 33       Brazil
## 34      Uruguay
## 35 NorthAmerica
## 36      Uruguay
## 37 SouthAmerica
## 38       Africa
## 39      Uruguay
## 40      Uruguay
## 41       Europe
```

```
## 42 NorthAmerica
## 43 SouthAmerica
## 44      Uruguay
## 45       Europe
## 46 NorthAmerica
## 47       Europe
## 48      Uruguay
## 49       Europe
```

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

```
## [1] 47  1
```

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

```
##
##      Africa      Asia      Brazil      Europe NorthAmerica SouthAmerica
##           6         3           1          13            3           11
##      Uruguay
##          12
```

```r
min(df$count)
```

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

```r
max(df$count)
```

```
## [1] 30
```

```r
# 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
```

```r
    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: 21"
```

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

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

```r
print(df)
```

```
##            label MRCA cluster count parent branch.length       date numeric_date
## 1   NODE_0003292 5337      90     2   5332       0.05234 2021-05-01  2021.331449
## 2   NODE_0004568 6466     572     5   6465       0.02861 2021-05-14  2021.366488
## 3   NODE_0003294 5338      92     2   5336       0.03650 2021-05-15  2021.367948
## 4   NODE_0003297 5340      95     6   5336       0.02167 2021-05-23  2021.389620
## 5   NODE_0003174 5308      76     4   5297       0.17550 2021-05-24  2021.394269
## 6   NODE_0004601 6495     579     5   6465       0.09553 2021-06-08  2021.433405
## 7   NODE_0003085 5235      40     9   5234       0.10765 2021-06-17  2021.459229
## 8   NODE_0003684 5760     265     5   5759       0.02676 2021-06-23  2021.474940
## 9   NODE_0003157 5293      65     5   5290       0.04103 2021-06-27  2021.486376
## 10  NODE_0003955 5454     159    18   5453       0.12267 2021-06-27  2021.486315
## 11  NODE_0003313 5351     101     2   5336       0.06543 2021-06-28  2021.490373
## 12  NODE_0004602 6539     593     6   6505       0.03939 2021-07-02  2021.499107
## 13  NODE_0004041 5632     210     3   5631       0.10060 2021-07-03  2021.503097
## 14  NODE_0004967 6429     553     2   6428       0.12285 2021-07-06  2021.511948
## 15  NODE_0003949 5448     157     6   5447       0.43855 2021-07-07  2021.513508
## 16  NODE_0005085 6571     601     2   6570       0.19990 2021-07-08  2021.516438
```

```
## 17 NODE_0003476 5780        283    5    5779        0.06914 2021-07-09  2021.519971
## 18 NODE_0004943 6419        548    3    6418        0.13712 2021-07-09  2021.519178
## 19 NODE_0000081 5247         43    8    5234        0.00236 2021-07-11  2021.524908
## 20 NODE_0001607 6550        595    2    6549        0.00235 2021-07-13  2021.530137
## 21 NODE_0003394 5969        381    4    5968        0.01505 2021-07-13  2021.530137
## 22 NODE_0003894 5424        150    4    5414        0.04240 2021-07-14  2021.533856
## 23 NODE_0003120 5259         48    2    5257        0.35181 2021-07-15  2021.535616
## 24 NODE_0004504 6356        521    3    6355        0.04150 2021-07-15  2021.535616
## 25 NODE_0005142 6281        469    4    6280        0.25109 2021-07-15  2021.535616
## 26 NODE_0005000 6437        561    3    6433        0.11709 2021-07-18  2021.545204
## 27 NODE_0003222 5380        115    2    5379        0.06480 2021-07-19  2021.546575
## 28 NODE_0003627 5827        296    2    5826        0.10458 2021-07-19  2021.546575
## 29 NODE_0003307 5348         97    2    5336        0.12711 2021-07-21  2021.552055
## 30 NODE_0003495 5794        292    2    5793        0.06171 2021-07-21  2021.552055
## 31 NODE_0004527 6333        507    5    6329        0.06827 2021-07-21  2021.552073
## 32 NODE_0003484 5787        285    4    5785        0.03175 2021-07-23  2021.557534
## 33 NODE_0004163 6085        418    4    6082        0.03805 2021-07-23  2021.558013
## 34 NODE_0001884 6775        711   13    6770        0.00243 2021-07-25  2021.563791
## 35 NODE_0001308 6294        480    5    6293        0.09123 2021-07-26  2021.565565
## 36 NODE_0001888 6777        707    2    6774        0.00784 2021-07-29  2021.573973
## 37 NODE_0004660 6519        589    5    6518        0.04286 2021-07-29  2021.573932
## 38 NODE_0004282 6121        423    2    6120        0.21243 2021-07-30  2021.576712
## 39 NODE_0001616 6555        597    2    6549        0.00245 2021-07-31  2021.579452
## 40 NODE_0001891 6779        709    4    6774        0.01096 2021-07-31  2021.579452
## 41 NODE_0004544 6317        498    9    6309        0.13377 2021-08-01  2021.582784
## 42 NODE_0003501 5797        294   30    5791        0.13816 2021-08-02  2021.585769
## 43 NODE_0003244 5330         86    2    5329        0.16947 2021-08-03  2021.588492
## 44 NODE_0001577 6525        591    2    6518        0.00249 2021-08-07  2021.598630
## 45 NODE_0004873 6738        694    7    6737        0.08791 2021-08-07  2021.598630
## 46 NODE_0005018 6299        483    2    6298        0.08400 2021-08-08  2021.601370
## 47 NODE_0004869 6764        700    4    6763        0.07476 2021-08-09  2021.602747
## 48 NODE_0004893 6798        716   13    6774        0.03275 2021-08-21  2021.636149
## 49 NODE_0004740 6687        684    2    6686        0.03823 2021-09-07  2021.682646
##          location
## 1          Africa
## 2     SouthAmerica
## 3          Africa
## 4          Africa
## 5          Africa
## 6     SouthAmerica
## 7          Africa
## 8          Europe
## 9            Asia
## 10           Asia
## 11         Africa
## 12    SouthAmerica
## 13    SouthAmerica
## 14         Europe
## 15           Asia
## 16         Europe
## 17    SouthAmerica
## 18         Europe
## 19         Africa
## 20    SouthAmerica
```

```
## 21        Europe
## 22        Europe
## 23        Africa
## 24 SouthAmerica
## 25        Africa
## 26        Europe
## 27        Europe
## 28        Europe
## 29        Africa
## 30 SouthAmerica
## 31 NorthAmerica
## 32 SouthAmerica
## 33        Brazil
## 34        Europe
## 35 NorthAmerica
## 36 SouthAmerica
## 37 SouthAmerica
## 38        Africa
## 39 SouthAmerica
## 40 SouthAmerica
## 41        Europe
## 42 NorthAmerica
## 43 SouthAmerica
## 44 SouthAmerica
## 45        Europe
## 46 NorthAmerica
## 47        Europe
## 48 SouthAmerica
## 49        Europe
```

```r
dim(as.data.frame(unique(df$parent))) # unique parent nodes outside Uruguay
```

```
## [1] 40  1
```

```r
t <- as.data.frame(table(df$location)) # number of introductions from each location
t <- t[order(t$Freq, decreasing = F),]
t
```

```
##           Var1 Freq
## 3        Brazil    1
## 2          Asia    3
## 5 NorthAmerica    4
## 1        Africa   11
## 4        Europe   14
## 6 SouthAmerica   16
```

```r
min(df$count) # minimum number of UY sequences in the a clusters
```

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

```r
max(df$count) # maximum number of UY sequences in the a clusters
```

```
## [1] 30
```

```r
p <- as.data.frame(table(df$location))
p <- p[order(p$Freq, decreasing = F),]
p
```

```
##            Var1 Freq
## 3        Brazil    1
## 2          Asia    3
## 5 NorthAmerica    4
## 1        Africa   11
## 4        Europe   14
## 6 SouthAmerica   16
```

```r
colnames(uy)[2] <- c("label")
meta <- read.csv("fx_final_metadata-dataset1.tsv",
    sep = "\t", header = T)
colnames(meta)[1] <- c("label")
k <- left_join(uy, meta, by = "label")
head(k)
```

```
##   cluster                                                        label
## 1      55 UY-CUY16-003792|EPI_ISL_NA|Uruguay|Montevideo|B.1.617.2|2021-07-05
## 2     446  UY-CUY17-003842|EPI_ISL_NA|Uruguay|Lavalleja|B.1.617.2|2021-07-05
## 3     553 UY-CUY17-003872|EPI_ISL_NA|Uruguay|Montevideo|B.1.617.2|2021-07-14
## 4     593        UY-CUY17-003887|EPI_ISL_NA|Uruguay|NA|B.1.617.2|2021-07-14
## 5     593        UY-CUY17-003888|EPI_ISL_NA|Uruguay|NA|B.1.617.2|2021-07-14
## 6     595        UY-CUY17-003889|EPI_ISL_NA|Uruguay|NA|B.1.617.2|2021-07-14
##         date  accession location        region country        city  subregion
## 1 2021-07-05 EPI_ISL_NA  Uruguay South America Uruguay Montevideo Montevideo
## 2 2021-07-05 EPI_ISL_NA  Uruguay South America Uruguay  Lavalleja  Lavalleja
## 3 2021-07-14 EPI_ISL_NA  Uruguay South America Uruguay Montevideo Montevideo
## 4 2021-07-14 EPI_ISL_NA  Uruguay South America Uruguay       <NA>    Uruguay
## 5 2021-07-14 EPI_ISL_NA  Uruguay South America Uruguay       <NA>    Uruguay
## 6 2021-07-14 EPI_ISL_NA  Uruguay South America Uruguay       <NA>    Uruguay
##   pango_lineage
## 1     B.1.617.2
## 2     B.1.617.2
## 3     B.1.617.2
## 4     B.1.617.2
## 5     B.1.617.2
## 6     B.1.617.2
```

```r
co2 <- k %>%
    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: 84 x 3
## # Groups:   cluster [84]
##    cluster status          count
##    <chr>   <chr>           <int>
## 1 101      not_published       2
## 2 115      not_published       2
## 3 117      not_published       1
## 4 148      not_published       1
## 5 150      not_published       4
## 6 152      not_published       1
## 7 154      not_published       1
## 8 157      not_published       6
## 9 159      not_published      18
## 10 177     not_published       1
## # i 74 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 Delta clade")
fig
```

Number of published and unpublished UY sequences in the Delta clade
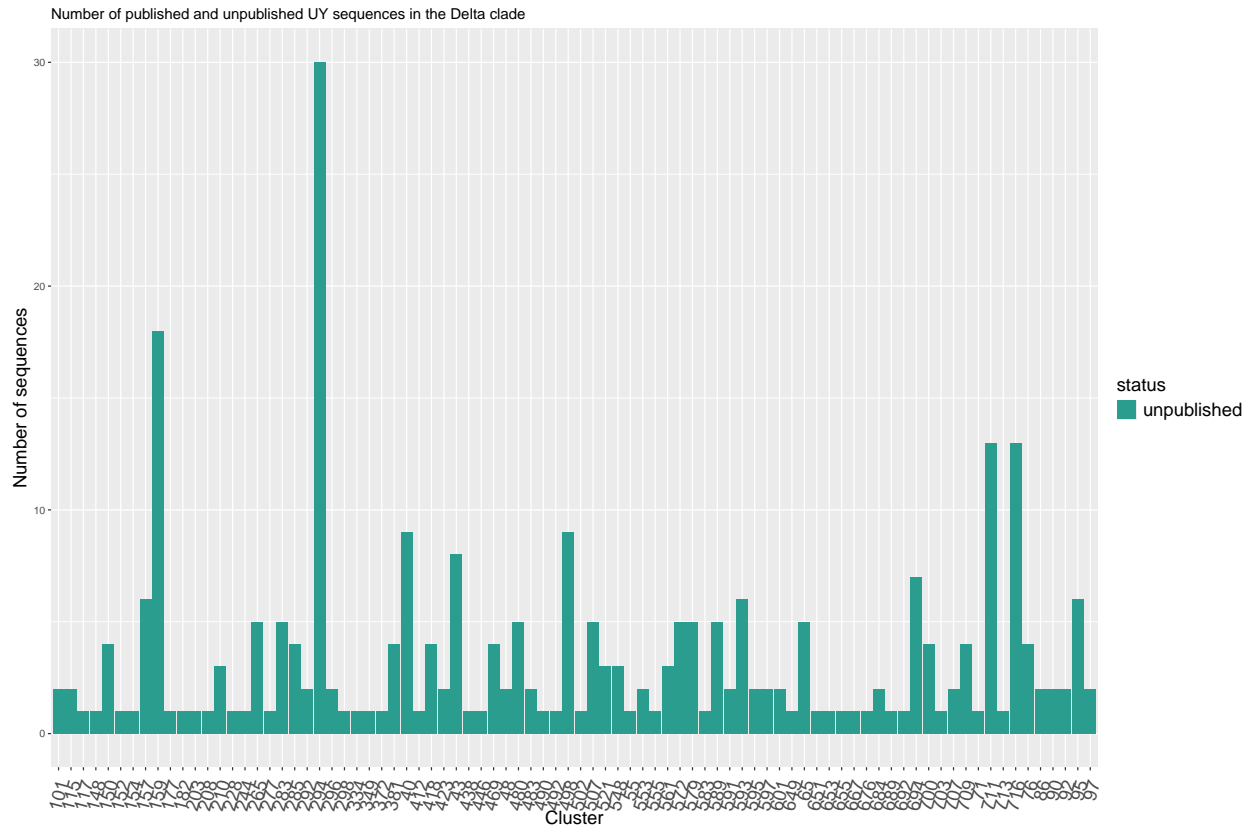

```r
tre <- read.beast("delta_subtree/joint_mugration_region_delta/annotated_tree.nexus")

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
```

```r
dim(cl)
```

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

```r
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   2084    1      1.31582
## 2   2086    2      0.04431
## 3   2087    3      0.11200
## 4   2088    4      0.00000
## 5   2088    5      0.01370
## 6   2090    6      0.00000
##                                                                 label location
## 1          BR-RS-LMM65724|EPI_ISL_6193430|Brazil|NA|None|2021-10-11   Brazil
## 2 ET-KRISP-CERI-K024753|EPI_ISL_4629090|Africa|NA|B.1.617.2|2021-05-25   Africa
## 3       UY-CUY19-004243|EPI_ISL_NA|Uruguay|NA|B.1.617.2|2021-07-28  Uruguay
## 4       UY-CUY19-004166|EPI_ISL_NA|Uruguay|NA|B.1.617.2|2021-07-22  Uruguay
## 5       UY-CUY19-004184|EPI_ISL_NA|Uruguay|NA|B.1.617.2|2021-07-27  Uruguay
## 6       UY-CUY17-003963|EPI_ISL_NA|Uruguay|NA|B.1.617.2|2021-07-13  Uruguay
##                      id       accession region subregion   lineage       date
## 1        BR-RS-LMM65724 EPI_ISL_6193430 Brazil        NA      None 2021-10-11
## 2 ET-KRISP-CERI-K024753 EPI_ISL_4629090 Africa        NA B.1.617.2 2021-05-25
## 3       UY-CUY19-004243     EPI_ISL_NA Uruguay        NA B.1.617.2 2021-07-28
## 4       UY-CUY19-004166     EPI_ISL_NA Uruguay        NA B.1.617.2 2021-07-22
## 5       UY-CUY19-004184     EPI_ISL_NA Uruguay        NA B.1.617.2 2021-07-27
## 6       UY-CUY17-003963     EPI_ISL_NA Uruguay        NA B.1.617.2 2021-07-13
```

```r
dim(t)
```

```
## [1] 3660   11
```

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

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

```
## [1] 280  12
```

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

```
## [1] 280  12
```

```r
meta <- t

cls <- c("#001219", "#005f73", "#0a9396",
    "#94d2bd", "#e9d8a6", "#ee9b00", "#ca6702",
```

```
    "#bb3e03", "#3a6ea5")

d <- ggtree(tre, aes(color = location), layout = "rectangular",
    size = 1, mrsd = "2021-12-28") + scale_color_manual(values = cls,
    labels = c("Africa", "Argentina", "Asia",
        "Brazil", "Europe", "North America",
        "Oceania", "South America", "Uruguay")) +
    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))
d
```

Region
- Africa
- Argentina
- Asia
- Brazil
- Europe
- North America
- Oceania
- South America
- Uruguay
- NA

2020–07   2021–01   2021–07   2022–01
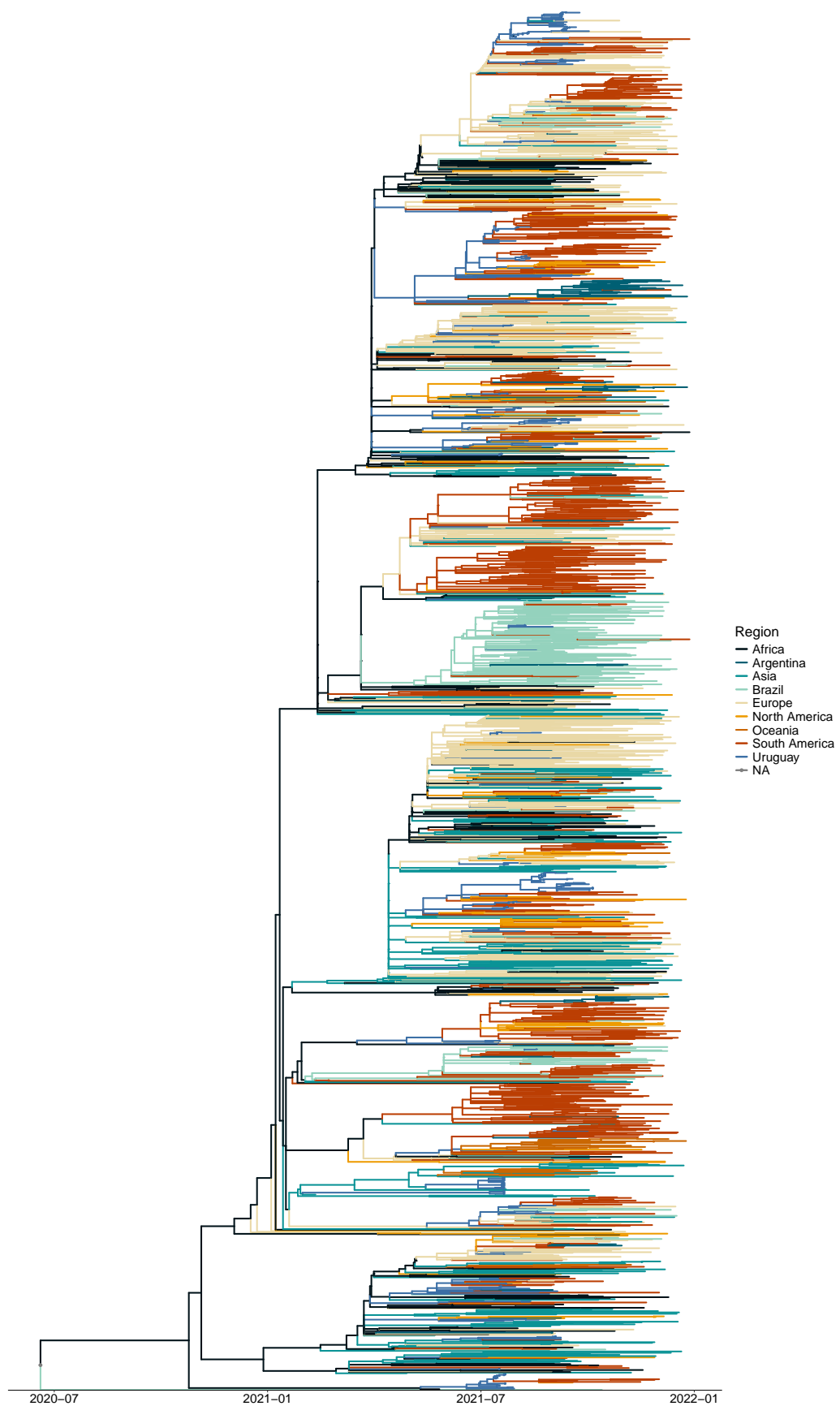
Conclusions

The tMRCA of the Delta sequences clusters from Uruguay dates from early May 2021. Based on unique parent nodes of the UY cluster, at least 47 introductions and re-introductions were detected. The main source of introductions are located in non-bordering countries from South America and Europe.

Create phylogeny

```r
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
```

```r
dim(cl)
```

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

```r
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-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
```

```r
dim(t)
```

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

```r
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)
```

```r
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"
```

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

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

```r
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))
```

```r
delta <- c("21J (Delta)", "21A (Delta)",
    "21I (Delta)")
delta <- filter(t, clade %in% delta)
delta_l <- delta$label

phy <- tre@phylo
delta_c <- findMRCA(phy, delta_l)
delta_f <- c("Clade")
dt <- as.data.frame(delta_c, delta_f)

g1 <- g + new_scale_fill()

cls <- c("#ae4d51")

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

Delta VOC

Region
- Africa
- Argentina
- Asia
- Brazil
- Europe
- NorthAmerica
- Oceania
- SouthAmerica
- Uruguay
- NA

```
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] vctrs_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
```