

# Estimating SARS-CoV-2 introductions into Uruguay

```
library(ape)
library(treeio)

## 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 v3.10.0 For help: https://yulab-smu.top/treedata-book/
##
## If you use the ggtree package suite in published research, please cite
## the appropriate paper(s):
##
## Guangchuang Yu, David Smith, Huachen Zhu, Yi Guan, Tommy Tsan-Yuk Lam.
## ggtree: an R package for visualization and annotation of phylogenetic
## trees with their covariates and other associated data. Methods in
## Ecology and Evolution. 2017, 8(1):28-36. doi:10.1111/2041-210X.12628
##
## Guangchuang Yu. Using ggtree to visualize data on tree-like structures.
## Current Protocols in Bioinformatics. 2020, 69:e96. doi:10.1002/cpbi.96
##
## 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
```

```
library(ggplot2)
library(phylostrate)
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
##
## 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: '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(tidyr)
```

```
##
## Attaching package: 'tidyr'

## The following object is masked from 'package:ggtree':
##
##     expand
```

```

library(stringr)
library(grid)
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v forcats 1.0.0      v readr 2.1.4
## v lubridate 1.9.3    v tibble 3.2.1
## v purrr 1.0.2

## -- Conflicts ----- tidyverse_conflicts() --
## x tidyr::expand() masks ggtree::expand()
## x dplyr::filter() masks tidytree::filter(), stats::filter()
## x dplyr::lag() masks stats::lag()
## x dplyr::where() masks ape::where()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(phangorn)
library(RColorBrewer)
library(data.table)

##
## Attaching package: 'data.table'
##
## The following objects are masked from 'package:lubridate':
##
##   hour, isoweek, mday, minute, month, quarter, second, wday, week,
##   yday, year
##
## The following object is masked from 'package:purrr':
##
##   transpose
##
## The following objects are masked from 'package:dplyr':
##
##   between, first, last

library(ggthemes)
library(ggpubr)

##
## Attaching package: 'ggpubr'
##
## The following object is masked from 'package:ggtree':
##
##   rotate
##
## The following object is masked from 'package:ape':
##
##   rotate

meta <- read.csv("fx_final_metadata-dataset1.tsv", sep = "\t", header = T)
dim(meta)

```

Find the oldest sampled sequence in Uruguay and the oldest sampled Gamma and Delta sequence in Uruguay.

```
## [1] 5173    9

uy <- meta[which(meta$location == "Uruguay"), ]
uy <- uy[order(uy$date, decreasing = F), ]
# Oldest UY sequence
uy[1, ]

##                                strain      date
## 1665 UY-Mdeo-1|EPI_ISL_444493|Uruguay|Montevideo|A.5|2020-03-13 2020-03-13
##           accession location      region country      city subregion
## 1665 EPI_ISL_444493  Uruguay South America Uruguay Montevideo Montevideo
##           pango_lineage
## 1665           A.5

gamma <- uy[grepl("P.1", uy$pango_lineage), ]
# Oldest Gamma sequence
gamma[1, ]

##                                strain      date
## 1036 UY-CUY16-003495|EPI_ISL_NA|Uruguay|NA|P.1.1|2021-02-19 2021-02-19
##           accession location      region country city subregion pango_lineage
## 1036 EPI_ISL_NA  Uruguay South America Uruguay <NA>  Uruguay      P.1.1

delta1 <- uy[grepl("B.1.617.2", uy$pango_lineage), ]
delta2 <- uy[grepl("AY.", uy$pango_lineage), ]
delta <- as.data.frame(rbind(delta1, delta2))
# Oldest Delta sequence
delta[1, ]

##                                strain
## 400 UY-CUY17-003842|EPI_ISL_NA|Uruguay|Lavalleya|B.1.617.2|2021-07-05
##           date accession location      region country      city subregion
## 400 2021-07-05 EPI_ISL_NA  Uruguay South America Uruguay Lavalleya Lavalleya
##           pango_lineage
## 400           B.1.617.2
```

## First SARS-CoV-2 introduction into Uruguay

```
tre <- read.nexus("treetime_out_joint/timetree.nexus")
t <- as_tibble(tre)
head(t)

## # A tibble: 6 x 4
##   parent node branch.length label
##   <int> <int>      <dbl> <chr>
## 1   5167     1      0.0641 CN-Hu-1|EPI_ISL_402125|Asia|NA|B|2019-12-26
## 2   5168     2      0.160  AU-VPRL014|EPI_ISL_455603|Oceania|NA|B.1|2020-01-30
## 3   5168     3      0.0641 CN-WH01|EPI_ISL_406798|Asia|NA|B|2019-12-26
## 4   5170     4      0.00546 AE-skmc-2200543130|EPI_ISL_582623|Asia|NA|B.53|202~
## 5   5170     5      0      AE-skmc-2200522146|EPI_ISL_582611|Asia|NA|B.53|202~
## 6   5171     6      0.0433 SG-1|EPI_ISL_406973|Asia|NA|B|2020-01-23

int <- c("UY-Mdeo-1|EPI_ISL_444493|Uruguay|Montevideo|A.5|2020-03-13",
"UY-UY-NYUMC864|EPI_ISL_457960|Uruguay|Montevideo|A.5|2020-03-28",
"UY-UY-NYUMC854|EPI_ISL_457950|Uruguay|Montevideo|A.5|2020-03-21",
"UY-UY-NYUMC844|EPI_ISL_457940|Uruguay|Montevideo|A.5|2020-03-17",
"UY-UY-NYUMC845|EPI_ISL_457941|Uruguay|Montevideo|A.5|2020-03-17",
```

```
"UY-UY-6|EPI_ISL_426481|Uruguay|Montevideo|A.5|2020-03-17", "UY-UY-5|EPI_ISL_426480|Uruguay|Montevideo|A.5|2020-03-17",
"UY-UY-NYUMC851|EPI_ISL_457947|Uruguay|Montevideo|A.5|2020-03-17",
"UY-UY-NYUMC852|EPI_ISL_457948|Uruguay|Montevideo|A.5|2020-03-17",
"UY-UY-NYUMC869|EPI_ISL_457965|Uruguay|Montevideo|A.5|2020-04-03",
"UY-UY-NYUMC846|EPI_ISL_457942|Uruguay|Montevideo|A.5|2020-03-17",
"UY-UY-3|EPI_ISL_426478|Uruguay|Montevideo|A.5|2020-03-17", "UY-UY-4|EPI_ISL_426479|Uruguay|Montevideo|A.5|2020-03-17",
"UY-UY-NYUMC855|EPI_ISL_457951|Uruguay|Montevideo|A.5|2020-03-21",
"UY-UY-NYUMC856|EPI_ISL_457952|Uruguay|Montevideo|A.5|2020-03-21",
"UY-UY-NYUMC847|EPI_ISL_457943|Uruguay|Montevideo|A.5|2020-03-17",
"UY-UY-1|EPI_ISL_426476|Uruguay|Montevideo|A.5|2020-03-16", "UY-UY-2|EPI_ISL_426477|Uruguay|Montevideo|A.5|2020-03-16",
"UY-UY-9|EPI_ISL_426584|Uruguay|Montevideo|A.5|2020-03-19", "UY-UY-NYUMC860|EPI_ISL_457956|Uruguay|Montevideo|A.5|2020-03-19",
"UY-UY-NYUMC862|EPI_ISL_457958|Uruguay|Montevideo|A.5|2020-03-25",
"UY-UY-NYUMC870|EPI_ISL_457966|Uruguay|Montevideo|A.5|2020-04-06",
"UY-UY-NYUMC930|EPI_ISL_480428|Uruguay|Montevideo|A.5|2020-03-26")
```

```
intro_cl <- MRCA(tre, int)
intro_cl_node <- t[grep(intro_cl, t$node), ]
intro_cl_node
```

```
## # A tibble: 1 x 4
##   parent node branch.length label
##   <int> <int>         <dbl> <chr>
## 1    5210  5211           0.0712 NODE_0003060
```

```
intro_node <- t[grep(intro_cl_node$label, t$label), ]
```

```
d <- read.csv("treetime_out_joint/dates.tsv", sep = "\t", header = T)
head(d)
```

```
##                               X.node      date numeric.date
## 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-WHO1|EPI_ISL_406798|Asia|NA|B|2019-12-26 2019-12-26  2019.984932
## 6                               NODE_0000002 2019-12-22  2019.974710
```

```
d <- d[order(d$date, decreasing = F), ]
tmrca <- filter(d, X.node %in% intro_node$label)
```

Results: The age of the root of the tree was located in 2019-11-06 (2019.846581). This date is similar to previous reports.

1. Nie et al (2020). Phylogenetic and phylodynamic analyses of SARS-CoV-2.
2. Duchene et al (2020). Temporal signal and the phylodynamic threshold of SARS-CoV-2.
3. Andersen et al (2020). The proximal origin of SARS-CoV-2.
4. Pekar et al (2021). Timing the SARS-CoV-2 index case in Hubei province.

The tMRCA of the node containing the oldest sequence in Uruguay is:

```
tmrca

##           X.node      date numeric.date
## 1 NODE_0003060 2020-02-19  2020.134762
```

```
tre <- read.beast("treetime_out_joint/fixed_mugration_region/annotated_tree.nexus")
t <- as.data.frame(as_tibble(tre))
t[grepl("Uruguay", t$Region), ]
```

Find the geographical origin of the first introduction to Uruguay

```
## [1] parent      node      branch.length label      location
## <0 rows> (or 0-length row.names)

intro_loc <- intro_cl_node$parent
intro_loc <- filter(t, node %in% intro_loc)
intro_loc
```

```
##   parent node branch.length      label      location
## 1   5186 5210      0.07646 NODE_0003056 SouthAmerica
```

Results: The tMRCA and origin (South America) of the first introduction of SARS-CoV-2 in Uruguay is similar to previously reported dates by Elizondo et al (2021) and reported in a pre-print by the group (Salazar et al.(2020)).

## Estimating the date of Gamma introduction

```
tre <- read.nexus("treetime_out_joint/timetree.nexus")
t <- as_tibble(tre)
head(t)
```

```
## # A tibble: 6 x 4
##   parent node branch.length label
##   <int> <int>      <dbl> <chr>
## 1   5167     1      0.0641 CN-Hu-1|EPI_ISL_402125|Asia|NA|B|2019-12-26
## 2   5168     2      0.160  AU-VPRL014|EPI_ISL_455603|Oceania|NA|B.1|2020-01-30
## 3   5168     3      0.0641 CN-WHO1|EPI_ISL_406798|Asia|NA|B|2019-12-26
## 4   5170     4      0.00546 AE-skmc-2200543130|EPI_ISL_582623|Asia|NA|B.53|202~
## 5   5170     5      0      AE-skmc-2200522146|EPI_ISL_582611|Asia|NA|B.53|202~
## 6   5171     6      0.0433 SG-1|EPI_ISL_406973|Asia|NA|B|2020-01-23
```

```
gamma <- c("UY-CUY6-001239|EPI_ISL_NA|Uruguay|Tacuarembó|P.1|2021-04-18",
"UY-CUY13-002967|EPI_ISL_NA|Uruguay|Tacuarembó|P.1.1|2021-06-11", "UY-CUY12-002788|EPI_ISL_NA|Uruguay|Tacuarembó|P.1.1|2021-06-11",
"UY-CUY13-002970|EPI_ISL_NA|Uruguay|Tacuarembó|P.1.1|2021-06-12", "UY-CUY12-002790|EPI_ISL_NA|Uruguay|Tacuarembó|P.1.1|2021-06-12",
"UY-CUY16-003565|EPI_ISL_NA|Uruguay|Rivera|P.1|2021-06-25", "UY-CUY12-002833|EPI_ISL_NA|Uruguay|Tacuarembó|P.1.1|2021-06-03",
"UY-CUY12-002787|EPI_ISL_NA|Uruguay|Tacuarembó|P.1|2021-06-03", "UY-CUY12-002767|EPI_ISL_NA|Uruguay|Tacuarembó|P.1.1|2021-06-04",
"UY-CUY12-002768|EPI_ISL_NA|Uruguay|Tacuarembó|P.1|2021-06-04", "UY-CUY16-003495|EPI_ISL_NA|Uruguay|Tacuarembó|P.1.1|2021-06-04",
"UY-CUY5-000729|EPI_ISL_NA|Uruguay|Tacuarembó|P.1|2021-04-10", "UY-CUY12-002834|EPI_ISL_NA|Uruguay|Tacuarembó|P.1.1|2021-04-10",
"UY-CUY8-002038|EPI_ISL_NA|Uruguay|NA|P.1|2021-05-10", "UY-CUY6-001343|EPI_ISL_NA|Uruguay|Montevideo|P.1|2021-04-20",
"UY-CUY6-001341|EPI_ISL_NA|Uruguay|Montevideo|P.1|2021-04-20")
```

```
gamma_cl <- MRCA(tre, gamma)
gamma_cl_node <- t[grepl(gamma_cl, t$node), ]
gamma_cl_node
```

```
## # A tibble: 1 x 4
##   parent node branch.length label
##   <int> <int>      <dbl> <chr>
## 1   8925 8926      0 NODE_0004764
```

```
gamma_node <- t[grepl(gamma_cl_node$label, t$label), ]

d <- read.csv("treetime_out_joint/dates.tsv", sep = "\t", header = T)
head(d)

##           X.node      date numeric.date
## 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

tmrca <- filter(d, X.node %in% gamma_node$label)
```

The tMRCA of the node containing the oldest Gamma VOC sequence in Uruguay is:

```
tmrca

##           X.node      date numeric.date
## 1 NODE_0004764 2021-01-30 2021.081183

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

gamma_loc <- gamma_cl_node$parent
gamma_loc <- filter(t, node %in% gamma_loc)
gamma_loc
```

Find the geographical origin of the first Gamma introduction to Uruguay

```
##   parent node branch.length      label location
## 1   8922 8925              0 NODE_0004763   Brazil
```

Results: The age of the node containing the oldest Gamma sequence is by the end of January of 2021 (2021-01-30) and the origin of this introduction is Brazil

## Estimating the age of the Delta clusters

Find the MRCA of the uruguayan cluster containing the oldest Delta sequence in the dataset.

```
tre <- read.nexus("treetime_out_joint/timetree.nexus")
t <- as_tibble(tre)
head(t)

## # A tibble: 6 x 4
##   parent node branch.length label
##   <int> <int>          <dbl> <chr>
## 1   5167     1      0.0641 CN-Hu-1|EPI_ISL_402125|Asia|NA|B|2019-12-26
## 2   5168     2      0.160  AU-VPRL014|EPI_ISL_455603|Oceania|NA|B.1|2020-01-30
## 3   5168     3      0.0641 CN-WH01|EPI_ISL_406798|Asia|NA|B|2019-12-26
## 4   5170     4      0.00546 AE-skmc-2200543130|EPI_ISL_582623|Asia|NA|B.53|202~
## 5   5170     5      0       AE-skmc-2200522146|EPI_ISL_582611|Asia|NA|B.53|202~
## 6   5171     6      0.0433 SG-1|EPI_ISL_406973|Asia|NA|B|2020-01-23
```

```

delta <- c("CO-BOY-INS-VG-9830|EPI_ISL_7314452|SouthAmerica|NA|AY.122|2021-10-17",
"EC-NIC-INSPI-101105|EPI_ISL_6507601|SouthAmerica|NA|AY.122|2021-10-08",
"CO-ARA-INS-VG-9759|EPI_ISL_7314395|SouthAmerica|NA|AY.122|2021-10-29",
"CO-ARA-IAVH-VG-10585|EPI_ISL_8000437|SouthAmerica|NA|AY.122|2021-10-29",
"CZ-NRL_16037|EPI_ISL_7126209|Europe|NA|AY.122|2021-10-28", "GE-Tb-SNGS2090|EPI_ISL_7876003|Asia|NA",
"LV-3446|EPI_ISL_8175176|Europe|NA|AY.122|2021-07-20", "UY-CUY17-003842|EPI_ISL_NA|Uruguay|Lavalleyj",

delta_cl <- MRCA(tre, delta)
delta_cl_node <- t[grep(delta_cl, t$node), ]
delta_cl_node

## # A tibble: 1 x 4
##   parent node branch.length label
##   <int> <int>         <dbl> <chr>
## 1   6186  6197           0.0236 NODE_0004419

delta_node <- t[grep(delta_cl_node$label, t$label), ]

d <- read.csv("treetime_out_joint/dates.tsv", sep = "\t", header = T)
head(d)

##                               X.node      date numeric.date
## 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

tmrca <- filter(d, X.node %in% delta_node$label)

```

The tMRCA of the node containing the oldest Delta VOC sequence in Uruguay is:

```

tmrca

##           X.node      date numeric.date
## 1 NODE_0004419 2021-05-17  2021.373415

```

```

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

delta_loc <- delta_cl_node$parent
delta_loc <- filter(t, node %in% delta_loc)
delta_loc

```

Find the geographical origin of the first Delta introduction to Uruguay

```

##   parent node branch.length      label location
## 1   6180 6186           0.02225 NODE_0004401   Europe

```

**Results:** The tMRCA of the tree was estimated in 2019-12-03 (2019.92082).

\textcolor{red}{The estimated time of the Delta variant introduction is May 2021 (2021-05-17) according to the tMRCA of the cluster containing the oldest Uruguayan Delta sequence. Also the node that contains the oldest UY delta sequence have origin in Europe.



## Joint tree visualization

```
tre <- read.nexus("treetime_out_joint/fixed_mugration_region/annotated_tree.nexus")
t <- as.tibble(tre)

## Warning: 'as.tibble()' was deprecated in tibble 2.0.0.
## i Please use 'as_tibble()' instead.
## i The signature and semantics have changed, see '?as_tibble'.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.

meta <- read.csv("treetime_out_joint/metadata_with_clade/datsaet_phyldynamics.tsv",
  sep = "\t", header = T)
colnames(meta)[1] <- c("label")

meta <- merge(t, meta, by = "label")
meta <- as.data.frame(meta[order(meta$date, decreasing = T), ])

p <- ggtree(tre, size = 0.1, color = "black", layout = "fan", mrsd = "2021-02-23") %<+%
  meta + theme(axis.text.x = element_blank(), axis.text.y = element_blank(),
  legend.text = element_text(size = 12), legend.title = element_text(size = 12)) +
  geom_rootpoint(position = "identity") + labs(colour = "Clado") + guides(colour = guide_legend(overr

## Scale for y is already present.
## Adding another scale for y, which will replace the existing scale.

clade <- select(meta, label, clade)
clade <- clade[!duplicated(clade$label), ]
row.names(clade) <- clade$label
clade$label <- NULL

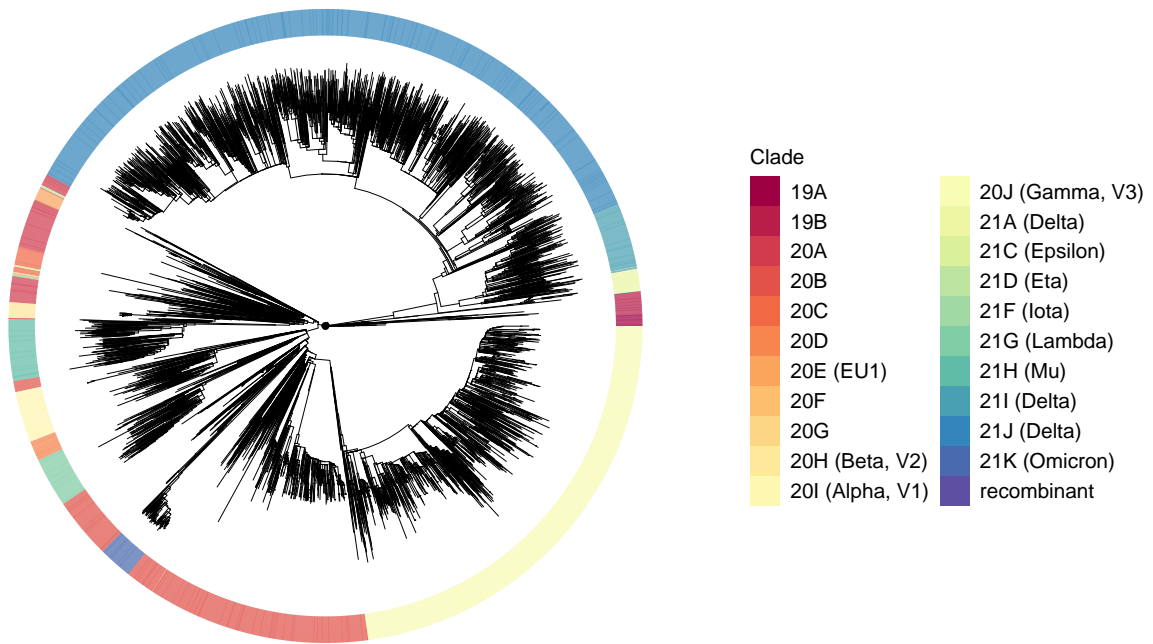
coul <- brewer.pal(11, "Spectral")
coul <- colorRampPalette(coul)(22)

p1 <- p + new_scale_fill()
t <- gheatmap(p1, clade, offset = 0.1, width = 0.1, colnames = F, color = F) +
  scale_fill_manual(values = coul, name = "Clade") + ggtitle("Time-scaled phylogeny using the joint i

## Scale for y is already present.
## Adding another scale for y, which will replace the existing scale.
## Scale for fill is already present.
## Adding another scale for fill, which will replace the existing scale.

t
```

Time-scaled phylogeny using the joint inference model of TreeTime



```
cls <- c("#ff0a47", "#0039a3", "#bfde26", "#00b2ea", "#60c888", "#8734d5",
        "#e50056", "#ff7b30", "#354f52", "#f94144", "#f3722c", "#f8961e", "#f9844a",
        "#f9c74f", "#90be6d", "#43aa8b", "#4d908e", "#577590", "#277da1", "#ea88a3",
        "#dd365b", "#a32b45", "#f2f1ff", "#bfbdd7", "#f4ecbf", "#fed92e", "#c8a221",
        "#e2547b", "#ad3b64")

tre <- read.beast("treetime_out_joint//fixed_mugration_region/annotated_tree.nexus")
t <- as.data.frame(as.tibble(tre))
meta <- meta[order(meta$date, decreasing = T), ]
head(meta)
```

```
##                                     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 date accession location region
## 1208 3309 0.09819 2021-12-28 EPI_ISL_8187354 Asia Asia
## 1939 1225 0.26301 2021-12-25 EPI_ISL_8207166 SouthAmerica South America
## 1940 2132 0.34359 2021-12-25 EPI_ISL_8207168 SouthAmerica South America
## 2025 3274 0.08493 2021-12-25 EPI_ISL_8189300 Asia Asia
## 2320 1538 0.65490 2021-12-25 EPI_ISL_8144256 Africa Africa
## 567 3279 0.04932 2021-12-24 EPI_ISL_8184739 Brazil South America
## country city subregion pango_lineage
## 1208 China Zhejiang / Hangzhou Asia BA.1
## 1939 French Guiana Saint Georges SouthAmerica AY.99.2
## 1940 French Guiana Macouria Tonate SouthAmerica AY.43
## 2025 Hong Kong Uruguay Asia BA.1
## 2320 Morocco Tanger Africa AY.98
## 567 Brazil Ceara / Fortaleza Brazil BA.1
## last_pango_lineage clade
## 1208 BA.1 21K (Omicron)
## 1939 AY.99.2 21J (Delta)
## 1940 AY.43 21J (Delta)
## 2025 BA.1.1 21K (Omicron)
## 2320 AY.98 21J (Delta)
## 567 BA.1.1 21K (Omicron)
```

```
d <- ggtree(tre, aes(color = location), size = 1, mrsd = "2021-12-28") +
  scale_colour_manual(values = cls) + theme_tree2() + geom_rootpoint(position = "identity") +
  xlim_tree(xlim = 2022) + scale_x_ggtree(labels = c("2020-01", "2020-07",
    "2021-01", "2021-07", "2022-01"), breaks = c(2020, 2020.5, 2021, 2021.5,
    2022))
```

```
f <- d %<+% meta + geom_tippoint(size = 1) + geom_rootpoint(position = "identity") +
  guides(colour = guide_legend("Region", override.aes = list(size = 2,
    shape = 19))) + theme(axis.text.x = element_text(size = 18), axis.text.y = element_blank(),
    legend.text = element_text(size = 20), legend.title = element_text(size = 20))
```

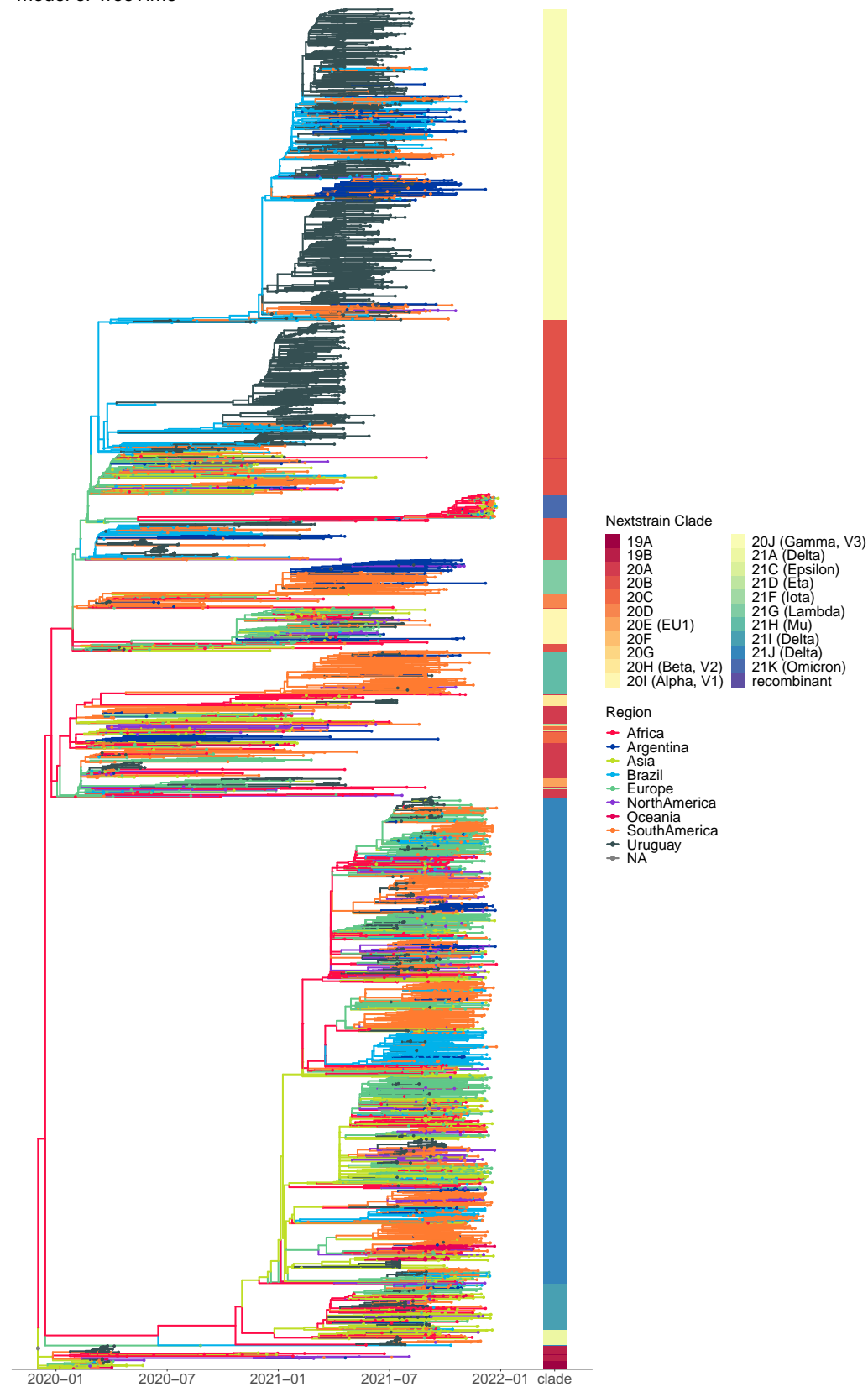
```
g2 <- f + new_scale_fill()
```

```
gg <- gheatmap(g2, clade, offset = 0.15, width = 0.05, colnames = FALSE,
  color = FALSE) + scale_y_continuous(expand = c(0, 0.4)) + scale_fill_manual(values = coul) +
  labs(fill = "Nextstrain Clade") + theme(axis.text.x = element_text(size = 18),
  axis.text.y = element_blank(), legend.text = element_text(size = 18),
  legend.title = element_text(size = 18), title = element_text(size = 20)) +
  scale_x_ggtree(labels = c("2020-01", "2020-07", "2021-01", "2021-07",
    "2022-01"), breaks = c(2020, 2020.5, 2021, 2021.5, 2022)) + ggtitle("Transition between geographies
  using the migration \n model of TreeTime")
```

```
## Scale for y is already present.
## Adding another scale for y, which will replace the existing scale.
## Scale for y is already present.
## Adding another scale for y, which will replace the existing scale.
## Scale for fill is already present.
## Adding another scale for fill, which will replace the existing scale.
```

```
## Scale for x is already present.  
## Adding another scale for x, which will replace the existing scale.  
gg
```

Transition between geographical states in the time-scaled phylogeny  
using the migration  
model of TreeTime



```
png("annotated_tree_joint.png", res = 800, height = 30, width = 25, units = "cm")
gg
dev.off()
```

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

## Geographical transitions (regions)

```
wt <- as.data.frame(read.table("treetime_out_joint/fixed_mugration_region/GTR_region_all.txt",
  sep = "\t", header = T))
head(wt)
```

```
##      X      A      B      C      D      E      F      G      H      I
## 1  A 0.0000 0.0217 0.3763 0.0507 0.3144 0.0812 0.0273 0.0361 0.0166
## 2  B 0.0739 0.0000 0.0480 0.1383 0.0603 0.1789 0.0596 0.1602 0.0851
## 3  C 0.6537 0.0245 0.0000 0.0524 0.5275 0.1210 0.1204 0.0489 0.0288
## 4  D 0.0733 0.0587 0.0436 0.0000 0.1105 0.0279 0.0195 0.1203 0.1493
## 5  E 0.4585 0.0259 0.4428 0.1117 0.0000 0.2053 0.0437 0.1071 0.0742
## 6  F 0.1367 0.0885 0.1172 0.0325 0.2369 0.0000 0.1025 0.1435 0.0267
```

```
wt <- mutate_all(wt, function(x) as.numeric(as.character(x)))
```

```
## Warning: There was 1 warning in 'mutate()'.
## i In argument: 'X = (function (x) ...'.
## Caused by warning:
## ! NAs introducidos por coerción
```

```
dim(wt)
```

```
## [1] 9 10
```

```
wt$X <- NULL
sum(wt)
```

```
## [1] 11.1117
```

```
abs <- wt/sum(wt) * 100
```

```
region <- as.data.frame(read.table("treetime_out_joint/fixed_mugration_region/GTR_region_regions_all.txt",
  sep = "\t", header = F))
region <- as.data.frame(region[-6, ])
```

```
split <- as.data.frame(str_split_fixed(region$`region[-6, ]`, ":", 2))
split$V2 <- gsub("SouthAmerica", "South America", split$V2)
split$V2 <- gsub("NorthAmerica", "North America", split$V2)
colnames(abs) <- split$V2
rownames(abs) <- split$V2
```

```
dt2 <- abs %>%
  rownames_to_column() %>%
  gather(colname, value, -rowname)
head(dt2)
```

```
##      rowname colname      value
```

```
## 1      Africa  Africa 0.0000000
## 2      Argentina Africa 0.6650648
## 3      Asia    Africa 5.8829882
## 4      Brazil  Africa 0.6596650
## 5      Europe  Africa 4.1262813
## 6 North America Africa 1.2302348

dt2$value <- gsub(",", ".", dt2$value)
dt2$value <- as.numeric(as.character(dt2$value))
dt2$value <- round(dt2$value, 1)
head(dt2)

##      rowname colname value
## 1      Africa  Africa   0.0
## 2      Argentina Africa   0.7
## 3      Asia    Africa   5.9
## 4      Brazil  Africa   0.7
## 5      Europe  Africa   4.1
## 6 North America Africa   1.2

region <- ggplot(dt2, aes(x = rowname, y = colname, fill = value)) + geom_tile(color = "white",
  lwd = 1.5, linetype = 1) + scale_fill_gradient2(low = "#075AFF", mid = "#FFFFCC",
  high = "#FF0000") + geom_text(aes(label = value), color = "black",
  size = 3) + coord_fixed() + theme(axis.text.x = element_text(angle = 90,
  vjust = 0.5, hjust = 1, size = 12), axis.text.y = element_text(size = 12),
  axis.title = element_text(size = 14), legend.text = element_text(size = 10),
  legend.title = element_blank(), panel.grid.major = element_blank(),
  panel.grid.minor = element_blank(), panel.border = element_blank(),
  panel.background = element_blank(), plot.title = element_text(size = 12)) +
  xlab("To") + ylab("From") + ggtitle("Transition rates (%) between regions")

abs

##      Africa Argentina      Asia      Brazil      Europe North America
## Africa      0.0000000 0.1952896 3.3865205 0.4562758 2.8294500      0.7307613
## Argentina 0.6650648 0.0000000 0.4319771 1.2446340 0.5426712      1.6100147
## Asia      5.8829882 0.2204883 0.0000000 0.4715750 4.7472484      1.0889423
## Brazil    0.6596650 0.5282720 0.3923792 0.0000000 0.9944473      0.2510867
## Europe     4.1262813 0.2330876 3.9849888 1.0052467 0.0000000      1.8476021
## North America 1.2302348 0.7964578 1.0547441 0.2924845 2.1319870      0.0000000
## Oceania    0.2861848 0.1835903 0.7262615 0.1412925 0.3140834      0.6380662
## South America 2.3290766 3.0418388 1.8161037 5.3817148 4.7436486      5.5077081
## Uruguay    0.8846531 1.3319294 0.8828532 5.5059082 2.7079565      0.8450552
##      Oceania South America      Uruguay
## Africa      0.2456870      0.3248828 0.1493921
## Argentina 0.5363716      1.4417236 0.7658594
## Asia      1.0835426      0.4400767 0.2591863
## Brazil    0.1754907      1.0826426 1.3436288
## Europe     0.3932792      0.9638489 0.6677646
## North America 0.9224511      1.2914316 0.2402873
## Oceania    0.0000000      0.2366875 0.1043945
## South America 1.4606226      0.0000000 2.2129827
## Uruguay    0.5327718      1.8242033 0.0000000
```

```
intro <- as.data.frame(t(abs[9, ]/sum(abs[9, ]) * 100))
intro$region <- row.names(intro)
row.names(intro) <- NULL
intro <- select(intro, region, Uruguay)
colnames(intro) <- c("Region", "rate")
intro$rate <- round(intro$rate, 1)
intro <- intro[-1, ]
intro
```

```
##           Region rate
## 2      Argentina  9.2
## 3           Asia  6.1
## 4        Brazil 37.9
## 5         Europe 18.7
## 6 North America  5.8
## 7         Oceania  3.7
## 8 South America 12.6
## 9          Uruguay  0.0
```

```
bar_i <- ggplot(intro, aes(x = reorder(Region, rate), y = rate, fill = Region)) +
  geom_bar(stat = "identity") + geom_text(aes(label = rate), hjust = 1,
    size = 3.5) + theme(axis.line = element_blank(), panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(), panel.border = element_blank(),
    panel.background = element_blank(), axis.text.x = element_text(hjust = 0.5),
    axis.title.x = element_text(size = 12), axis.title.y = element_text(size = 12,
      vjust = 1.5), axis.text.y = element_text(size = 10), legend.position = "none",
    plot.title = element_text(size = 10)) + scale_fill_manual(values = c(Brazil = "#fcb186",
    Europe = "#f3ffbd", `South America` = "#ec9393", Argentina = "#e6d186",
    Asia = "#eca7ba", Africa = "#d0d9b6", `North America` = "#dca76c",
    Oceania = "#ffd4d7")) + ylab("") + xlab("") + coord_flip() + ggtitle("To Uruguay")
```

```
dif <- as.data.frame(abs[, 9]/sum(abs[, 9]) * 100)
row.names(dif) <- row.names(abs)
dif$Region <- row.names(dif)
row.names(dif) <- NULL
colnames(dif) <- c("rate", "Region")
dif <- select(dif, Region, rate)
dif <- dif[order(dif$rate, decreasing = F), ]
dif <- dif[-1, ]
dif$rate <- round(dif$rate, 1)
```

```
bar_e <- ggplot(dif, aes(x = reorder(Region, rate), y = rate, fill = Region)) +
  geom_bar(stat = "identity") + geom_text(aes(label = rate), hjust = 1,
    size = 3.5) + theme(axis.line = element_blank(), panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(), panel.border = element_blank(),
    panel.background = element_blank(), axis.text.x = element_text(hjust = 0.5),
    axis.title.x = element_text(size = 12), axis.title.y = element_text(size = 12,
      vjust = 1.5), axis.text.y = element_text(size = 10), legend.position = "none",
    plot.title = element_text(size = 10)) + scale_fill_manual(values = c(Brazil = "#fcb186",
    Europe = "#f3ffbd", `South America` = "#ec9393", Argentina = "#e6d186",
    Asia = "#eca7ba", Africa = "#d0d9b6", `North America` = "#dca76c",
```



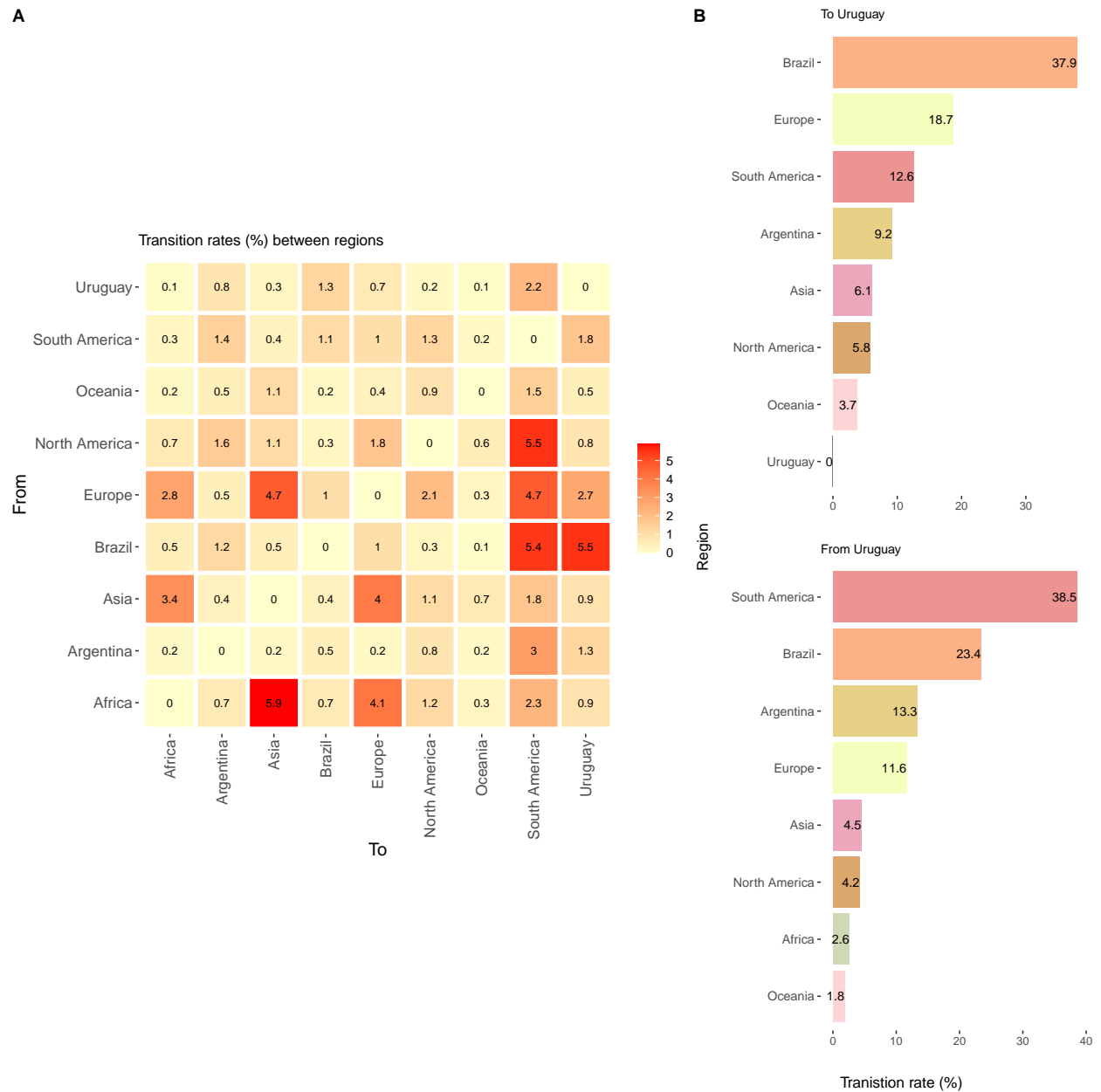
```

Oceania = "#ffd4d7")) + ylab("") + xlab("") + coord_flip() + ggtitle("From Uruguay")

figb <- ggarrange(bar_i, bar_e, nrow = 2)
figb <- annotate_figure(figb, left = textGrob("Region", rot = 90, vjust = 1,
  gp = gpar(cex = 1)), bottom = textGrob("Transition rate (%)", gp = gpar(cex = 1)))

fig1 <- ggarrange(region, figb, ncol = 2, labels = c("A", "B"), heights = c(1,
  0.6), widths = c(1, 0.6))
plot(fig1)

```



```

png("transition_region.png", res = 800, height = 30, width = 55, units = "cm")
fig1

```

```
dev.off()
```

```
## pdf
```

```
## 2
```