

Supplementary 3

Nextclade and PANGO lineages from genomes

```
library(dplyr)

##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(ggplot2)
library(stringr)

cl <- read.table("classif_s1_s28_next_pango.tsv", header = T, sep = "\t")
head(cl)

##           seqName           clade      id lineage_t03
## 1 CUY1-000001|B.1.1.28|2021-03-04      20B CUY1-000001      B.1.1.28
## 2      CUY1-000006|P.2|2021-03-04      20B CUY1-000006          P.2
## 3 CUY1-000008|B.1.1.28|2021-03-04      20B CUY1-000008      B.1.1.28
## 4      CUY1-000011|P.1|2021-03-04 20J (Gamma, V3) CUY1-000011          P.1
## 5      CUY1-000020|P.1|2021-03-04 20J (Gamma, V3) CUY1-000020          P.1
## 6 CUY1-000021|B.1.1.28|2021-03-04      20B CUY1-000021      B.1.1.28
##      date_t03 lineage
## 1 2021-03-04      P.7
## 2 2021-03-04      P.2
## 3 2021-03-04      P.6
## 4 2021-03-04      P.1
## 5 2021-03-04      P.1
## 6 2021-03-04      P.6

colnames(cl)[3] <- c("Identificador.de.Consorcio")

ck <- read.table("summary_with_depth.tsv", header = T, sep = "\t")
head(ck)

##      Sample_ID Diagnostic_Ct      Date Country Departament      Sex Age
## 1      CUY1-000139          14,0 2020-12-01 Uruguay  Tacuarembó      Male
## 2 REICUY1-000141          13,0 2020-12-01 Uruguay  Tacuarembó      Male
## 3 REICUY1-000145          18,2 2020-12-18 Uruguay  Tacuarembó      Male
## 4 REICUY1-000148          18,0 2020-12-22 Uruguay  Tacuarembó      Male
## 5      CUY1-000152          16,3 2020-12-24 Uruguay  Tacuarembó Female
## 6 REICUY1-000166          17,3 2021-01-11 Uruguay  Tacuarembó      Male
##      Epidemiologycal_Link PANGO_lineage Nextstrain_clade Depth.X.
## 1                          P.2                20B      842,6
```

```
## 2          B.1.1.33          20B      528,0
## 3          B.1.1.33          20B      110,6
## 4          P.2              20B      132,5
## 5          B.1.1.33          20B      733,2
## 6          B.1.1            20B      106,4
```

```
colnames(ck)[1] <- c("Identificador.de.Consorcio")
```

```
merged <- merge(cl, ck, by = "Identificador.de.Consorcio")
dim(merged)
```

```
## [1] 1049    16
```

```
head(merged)
```

```
##   Identificador.de.Consorcio          seqName          clade
## 1          CUY1-000006          CUY1-000006|P.2|2021-03-04          20B
## 2          CUY1-000061 CUY1-000061|B.1.1.28|2021-03-11          20B
## 3          CUY1-000139          CUY1-000139|P.2|2020-12-01          20B
## 4          CUY1-000152 CUY1-000152|B.1.1.33|2020-12-24          20B
## 5          CUY1-000172          CUY1-000172|P.2|2021-01-21          20B
## 6          CUY10-002304          CUY10-002304|P.1|2021-05-19 20J (Gamma, V3)
##   lineage_t03   date_t03   lineage Diagnostic_Ct   Date Country Departament
## 1          P.2 2021-03-04          P.2          18,0 2021-03-04 Uruguay      Artigas
## 2    B.1.1.28 2021-03-11          P.7          16,4 2021-03-11 Uruguay      Rocha
## 3          P.2 2020-12-01          P.2          14,0 2020-12-01 Uruguay  Tacuarembó
## 4    B.1.1.33 2020-12-24 B.1.1.33          16,3 2020-12-24 Uruguay  Tacuarembó
## 5          P.2 2021-01-21          P.2          13,4 2021-01-21 Uruguay  Tacuarembó
## 6          P.1 2021-05-19          P.1          21,9 2021-05-19 Uruguay      Rocha
##   Sex Age Epidemiological_Link PANGO_lineage Nextstrain_clade Depth.X.
## 1  Male  38                      P.2          20B      992,1
## 2  Male  45                      P.7          20B      956,2
## 3  Male                      P.2          20B      842,6
## 4 Female          B.1.1.33          20B      733,2
## 5 Female                      P.2          20B      779,4
## 6  Male          P.1 20J (Gamma, V3)          126,5
```

```
merged$check_clade <- merged$clade == merged$Nextstrain_clade
unique(merged$check_clade)
```

```
## [1] TRUE FALSE
```

```
merged[which(merged$check_clade == "FALSE"),]
```

```
##   Identificador.de.Consorcio          seqName clade lineage_t03
## 590          CUY19-004104 CUY19-004104|P.1.1|2021-07-21 20B      P.1.1
##   date_t03 lineage Diagnostic_Ct   Date Country Departament Sex Age
## 590 2021-07-21          P.1          22,0 2021-07-21 Uruguay      Artigas Male  53
##   Epidemiological_Link PANGO_lineage Nextstrain_clade Depth.X. check_clade
## 590          P.1 20J (Gamma, V3)          475,9          FALSE
```

```
merged$check_pango <- merged$lineage== merged$PANGO_lineage
unique(merged$check_pango)
```

```
## [1] TRUE FALSE
```

```
merged[which(merged$check_pango == "FALSE"),]
```

##	Identificador.de.Consorcio	seqName	clade
## 652	CUY20-004404	CUY20-004404 P.1.1 none	20J (Gamma, V3)
## 653	CUY20-004405	CUY20-004405 P.1.1 none	20J (Gamma, V3)
## 892	CUY6-000798	CUY6-000798 P.1 2021-04-18	20J (Gamma, V3)
## 897	CUY6-000823	CUY6-000823 P.1.1 2021-04-19	20J (Gamma, V3)
## 901	CUY6-000838	CUY6-000838 P.1 2021-04-19	20J (Gamma, V3)
## 904	CUY6-000855	CUY6-000855 P.1.1 2021-04-20	20J (Gamma, V3)
## 926	CUY6-000945	CUY6-000945 P.1 2021-04-19	20J (Gamma, V3)
## 927	CUY6-000951	CUY6-000951 P.1.1 2021-04-20	20J (Gamma, V3)
## 946	CUY6-001057	CUY6-001057 P.1 2021-04-18	20J (Gamma, V3)
## 956	CUY6-001118	CUY6-001118 P.1 2021-04-20	20J (Gamma, V3)
## 976	CUY6-001205	CUY6-001205 P.1 2021-04-18	20J (Gamma, V3)
## 1020	CUY7-001410	CUY7-001410 P.1 2021-04-12	20J (Gamma, V3)
## 1022	CUY7-001417	CUY7-001417 P.1 2021-04-12	20J (Gamma, V3)
## 1026	CUY7-001439	CUY7-001439 P.1 2021-04-12	20J (Gamma, V3)
## 1028	CUY7-001441	CUY7-001441 P.1 2021-04-12	20J (Gamma, V3)
## 1031	CUY7-001445	CUY7-001445 P.1 2021-04-12	20J (Gamma, V3)
## 1032	CUY7-001481	CUY7-001481 P.1 2021-04-22	20J (Gamma, V3)
## 1048	REICUY6-000995	REICUY6-000995 P.1 2021-04-19	20J (Gamma, V3)
## 1049	REICUY6-001193	REICUY6-001193 P.1 2021-04-20	20J (Gamma, V3)

##	lineage_t03	date_t03	lineage	Diagnostic_Ct	Date	Country
## 652	P.1.1	none	P.1.7		2021-08-04	Uruguay
## 653	P.1.1	none	P.1.7		2021-08-04	Uruguay
## 892	P.1	2021-04-18	P.1	15,9	2021-04-18	Uruguay
## 897	P.1.1	2021-04-19	P.1.10	25,4	2021-04-19	Uruguay
## 901	P.1	2021-04-19	P.1.10	18,7	2021-04-19	Uruguay
## 904	P.1.1	2021-04-20	P.1	17,3	2021-04-20	Uruguay
## 926	P.1	2021-04-19	P.1	23,9	2021-04-19	Uruguay
## 927	P.1.1	2021-04-20	P.1	25,6	2021-04-20	Uruguay
## 946	P.1	2021-04-18	P.1.10	17,4	2021-04-18	Uruguay
## 956	P.1	2021-04-20	P.1	18,7	2021-04-20	Uruguay
## 976	P.1	2021-04-18	P.1	23,9	2021-04-18	Uruguay
## 1020	P.1	2021-04-12	P.1	14,9	2021-04-12	Uruguay
## 1022	P.1	2021-04-12	P.1	14,7	2021-04-12	Uruguay
## 1026	P.1	2021-04-12	P.1	17,1	2021-04-12	Uruguay
## 1028	P.1	2021-04-12	P.1	16,2	2021-04-12	Uruguay
## 1031	P.1	2021-04-12	P.1	14,0	2021-04-12	Uruguay
## 1032	P.1	2021-04-22	P.1	15,1	2021-04-22	Uruguay
## 1048	P.1	2021-04-19	P.1	17,8	2021-04-19	Uruguay
## 1049	P.1	2021-04-20	P.1	19,7	2021-04-20	Uruguay

##	Departament	Sex	Age	Epidemiologcal_Link	PANGO_lineage	Nextstrain_clade
## 652	unknown				P.1	20J (Gamma, V3)
## 653	unknown				P.1	20J (Gamma, V3)
## 892	Montevideo	Female			P.1.12	20J (Gamma, V3)
## 897	Montevideo	Male			P.1	20J (Gamma, V3)
## 901	Montevideo	Male			P.1	20J (Gamma, V3)
## 904	Montevideo	Male			P.1.11	20J (Gamma, V3)
## 926	Artigas	Male	59		P.1.12	20J (Gamma, V3)
## 927	Canelones	Male	2		P.1.12	20J (Gamma, V3)
## 946	Canelones	Female	24		P.1	20J (Gamma, V3)
## 956	Canelones	Male	63		P.1.12	20J (Gamma, V3)
## 976	Tacuarembó	Female	45		P.1.12	20J (Gamma, V3)
## 1020	unknown	Female			P.1.12	20J (Gamma, V3)
## 1022	unknown	Male			P.1.12	20J (Gamma, V3)

```
## 1026      unknown Female          P.1.12  20J (Gamma, V3)
## 1028      unknown Female          P.1.12  20J (Gamma, V3)
## 1031      unknown Female          P.1.12  20J (Gamma, V3)
## 1032      Rocha Female            P.1.12  20J (Gamma, V3)
## 1048      Canelones Male 34        P.1.12  20J (Gamma, V3)
## 1049      Canelones Male 9         P.1.12  20J (Gamma, V3)
##      Depth.X. check_clade check_pango
## 652      477,3      TRUE      FALSE
## 653      400,9      TRUE      FALSE
## 892      495,0      TRUE      FALSE
## 897      138,8      TRUE      FALSE
## 901      355,3      TRUE      FALSE
## 904      260,0      TRUE      FALSE
## 926      526,0      TRUE      FALSE
## 927      665,0      TRUE      FALSE
## 946      725,3      TRUE      FALSE
## 956      522,2      TRUE      FALSE
## 976      564,1      TRUE      FALSE
## 1020     432,7      TRUE      FALSE
## 1022     463,9      TRUE      FALSE
## 1026     323,0      TRUE      FALSE
## 1028     401,3      TRUE      FALSE
## 1031     392,7      TRUE      FALSE
## 1032     304,7      TRUE      FALSE
## 1048     381,6      TRUE      FALSE
## 1049     307,7      TRUE      FALSE
```

#usar cl para clado y linaje ya que es lo mas actualizado

```
ck$PANGO_lineage <- NULL
```

```
cl[,4:5] <- NULL
```

```
merged <- merge(cl, ck, by = "Identificador.de.Consorcio")
```

```
head(merged)
```

```
##      Identificador.de.Consorcio      seqName      clade
## 1      CUY1-000006      CUY1-000006|P.2|2021-03-04      20B
## 2      CUY1-000061 CUY1-000061|B.1.1.28|2021-03-11      20B
## 3      CUY1-000139      CUY1-000139|P.2|2020-12-01      20B
## 4      CUY1-000152 CUY1-000152|B.1.1.33|2020-12-24      20B
## 5      CUY1-000172      CUY1-000172|P.2|2021-01-21      20B
## 6      CUY10-002304      CUY10-002304|P.1|2021-05-19 20J (Gamma, V3)
##      lineage Diagnostic_Ct      Date Country Departament      Sex Age
## 1      P.2      18,0 2021-03-04 Uruguay      Artigas      Male 38
## 2      P.7      16,4 2021-03-11 Uruguay      Rocha      Male 45
## 3      P.2      14,0 2020-12-01 Uruguay      Tacuarembó      Male
## 4 B.1.1.33      16,3 2020-12-24 Uruguay      Tacuarembó Female
## 5      P.2      13,4 2021-01-21 Uruguay      Tacuarembó Female
## 6      P.1      21,9 2021-05-19 Uruguay      Rocha      Male
##      Epidemiologcal_Link Nextstrain_clade Depth.X.
## 1      20B      992,1
## 2      20B      956,2
## 3      20B      842,6
## 4      20B      733,2
## 5      20B      779,4
## 6      20J (Gamma, V3)      126,5
```

```

clade <- as.data.frame(table(merged$Nextstrain_clade))
clade <- clade[order(clade$Freq, decreasing = T),]
clade$ptc <- clade$Freq/sum(clade$Freq)*100

c20j <- merged[which(merged$Nextstrain_clade == "20J (Gamma, V3)",)]
c20jL <- as.data.frame(table(c20j$lineage))
c20jL$pct <- c20jL$Freq/sum(c20jL$Freq)*100

c21j <- merged[which(merged$clade == "21J (Delta)",)]
c21jL <- as.data.frame(table(c21j$lineage))
c21jL$pct <- c21jL$Freq/sum(c21jL$Freq)*100
c21jL <- c21jL[order(c21jL$Freq, decreasing = T),]

#1049 with Ns < 15% available for analyzing, 118 of these samples have no date associated. Probably tra
list <- merged %>%
  group_by(Date, lineage, clade) %>%
  tally() %>%
  ungroup()
list <- as.data.frame(list)
dim(list)

## [1] 344    4

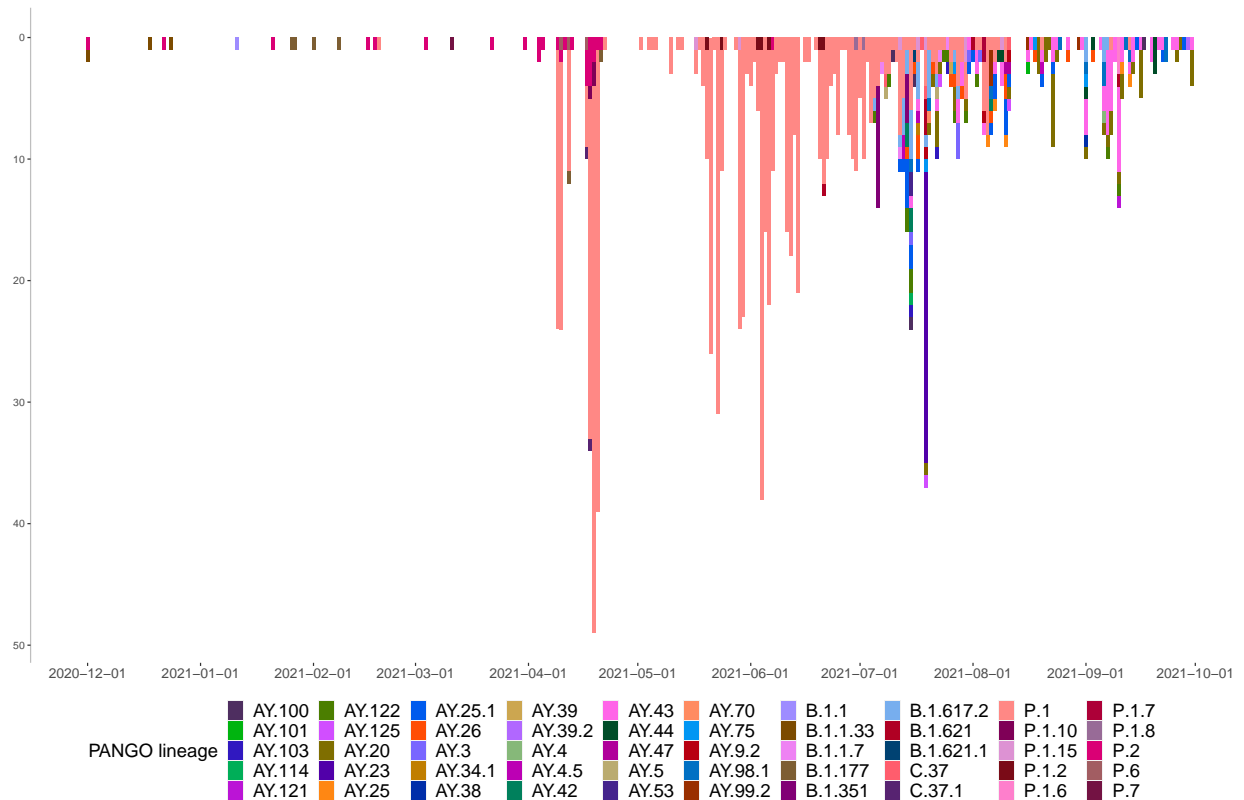
head(list)

##           Date lineage clade n
## 1 2020-12-01 B.1.1.33  20B  1
## 2 2020-12-01      P.2  20B  1
## 3 2020-12-18 B.1.1.33  20B  1
## 4 2020-12-22      P.2  20B  1
## 5 2020-12-24 B.1.1.33  20B  1
## 6 2021-01-11   B.1.1  20B  1

list$Date <- as.Date(list$Date)

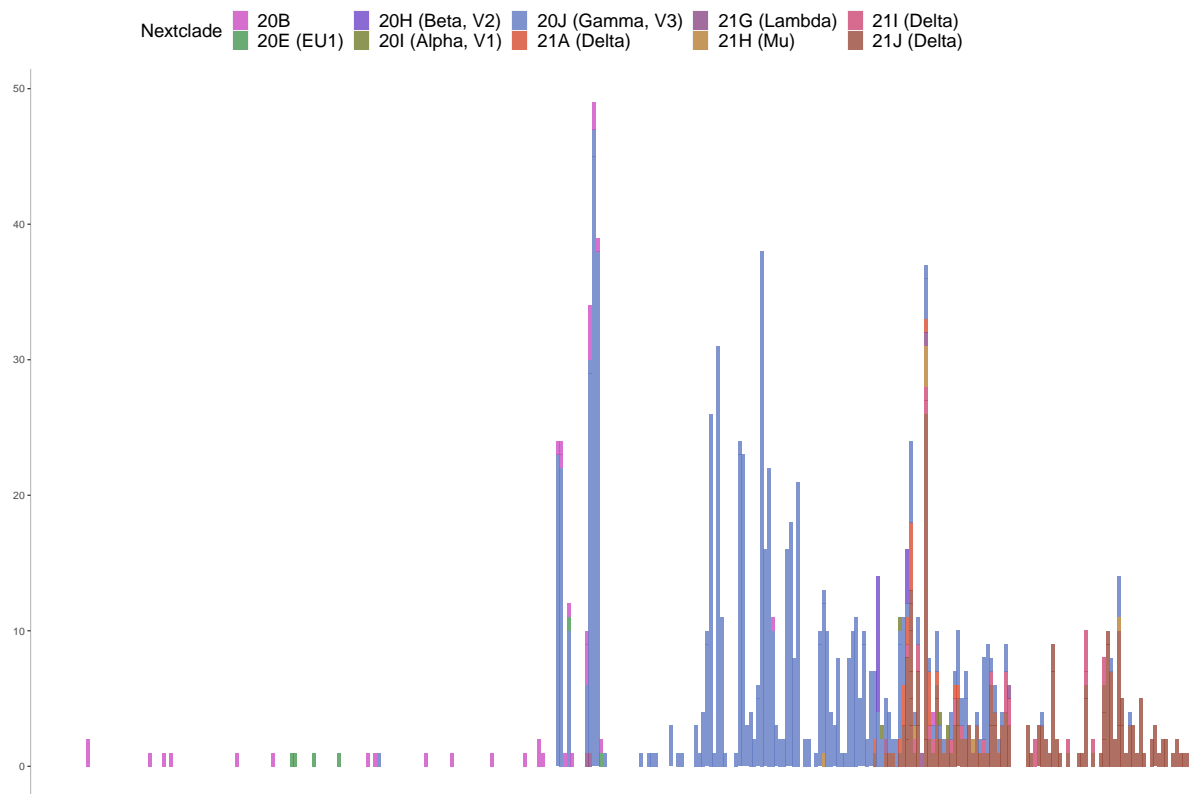
s3 <- ggplot(list, aes(fill = lineage, y = n, x = Date)) + geom_bar(position = "stack",
  stat = "identity", alpha = 1) + scale_fill_manual(values = c("#4e2e60", "#00b40f",
  "#3018c0", "#00ae58", "#bb14d6", "#4b7e00", "#d14eff", "#7f6e00", "#5100a9",
  "#ff8713", "#005bed", "#ff4d04", "#7966ff", "#c07e00", "#002ea9", "#cca650",
  "#b167ff", "#86b879", "#bc00b4", "#00805c", "#ff65e8", "#004c28", "#b0009a",
  "#baab70", "#45248d", "#ff8b61", "#0196f3", "#b80011", "#0170c3", "#992f00",
  "#9e8cff", "#7c4b00", "#ee82f3", "#7d5e33", "#810075", "#77aeec", "#b00023",
  "#004272", "#ff5d6d", "#572371", "#ff8885", "#7f0057", "#dd92d3", "#7a0b17",
  "#fe7fcc", "#ad0038", "#986b97", "#da0075", "#a25d62", "#731243")) + theme(axis.line = element_line
  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, vjust = 1,
    size = 12), axis.line.x = element_blank(), axis.title.x = element_blank(),
  axis.title.y = element_blank(), legend.text = element_text(size = 16), legend.title = element_text(
  legend.position = "bottom", legend.box.just = "center", legend.key.width = unit(1,
    "lines")) + scale_x_date(date_breaks = "1 months") + ggtitle("") + guides(fill = guide_legend(t
  nrow = 5, ncol = 11)) + scale_y_continuous(trans = "reverse", breaks = seq(0,
  80, by = 10)) + ylab("Log scale number of genomes") + xlab("Date")
s3

```

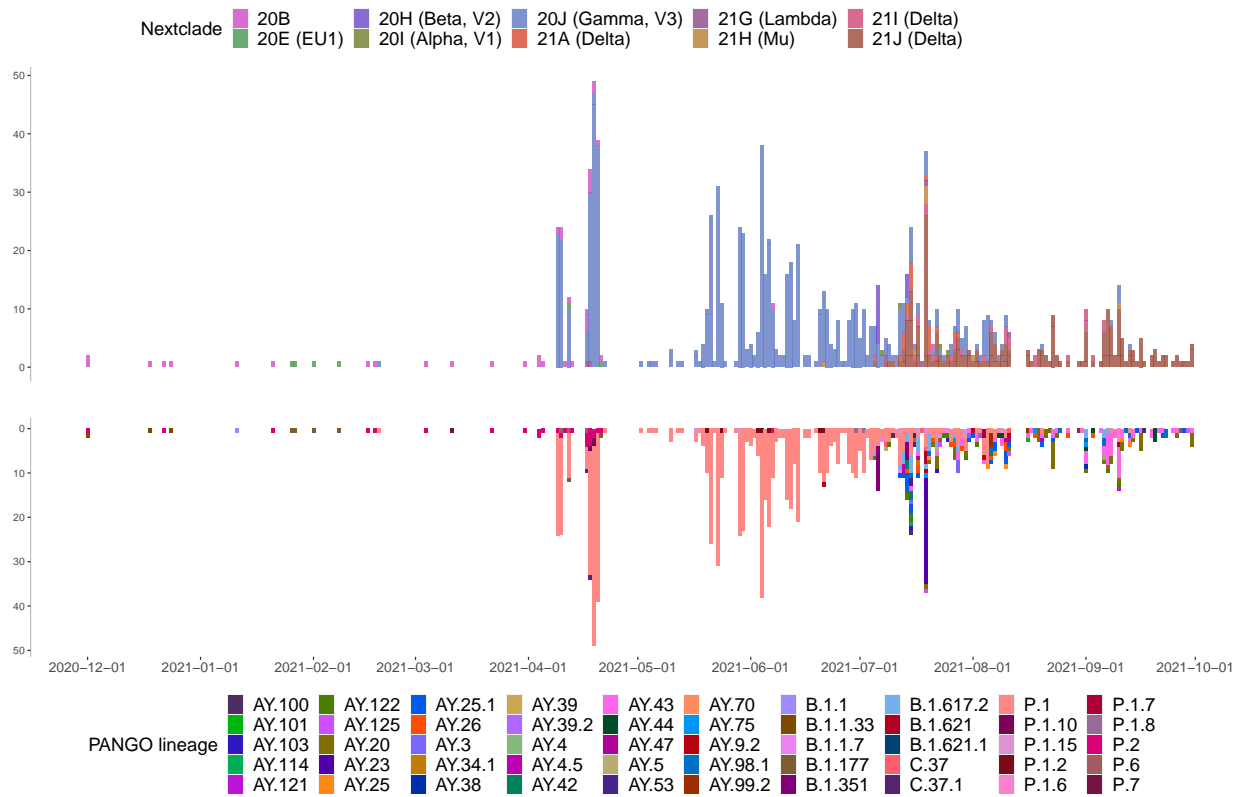


```
s4 <- ggplot(list, aes(fill = clade, y = n, x = Date)) + geom_bar(position = "stack",
  stat = "identity", alpha = 0.8) + scale_fill_manual(values = c("#cf4ec5", "#489954",
    "#7349cd", "#747f2e", "#617bc4", "#db4e33", "#904b8a", "#bc8136", "#d04a77",
    "#9c4c3c")) + theme(axis.line = element_line(colour = "grey"), panel.grid.major = element_blank(),
  panel.grid.minor = element_blank(), panel.border = element_blank(), panel.background = element_blank(),
  axis.text.x = element_blank(), axis.line.x = element_blank(), axis.title.x = element_blank(),
  axis.ticks.x = element_blank(), axis.title.y = element_blank(), legend.text = element_text(size = 10),
  legend.title = element_text(size = 16), legend.position = "top", legend.box.just = "center",
  legend.key.width = unit(1, "lines")) + scale_x_date(date_breaks = "2 months") +
  ggtitle("") + scale_y_continuous(breaks = seq(0, 80, by = 10)) + guides(fill = guide_legend(title = "PANGO lineage",
    nrow = 2, ncol = 10)) + ylab("Log scale number of genomes") + xlab("Date")
```

s4



```
library(ggpubr)
sup <- ggarrange(s4, s3, nrow = 2, vjust = 2, widths = c(0.5, 0.5), heights = c(0.5,
0.5))
sup
```

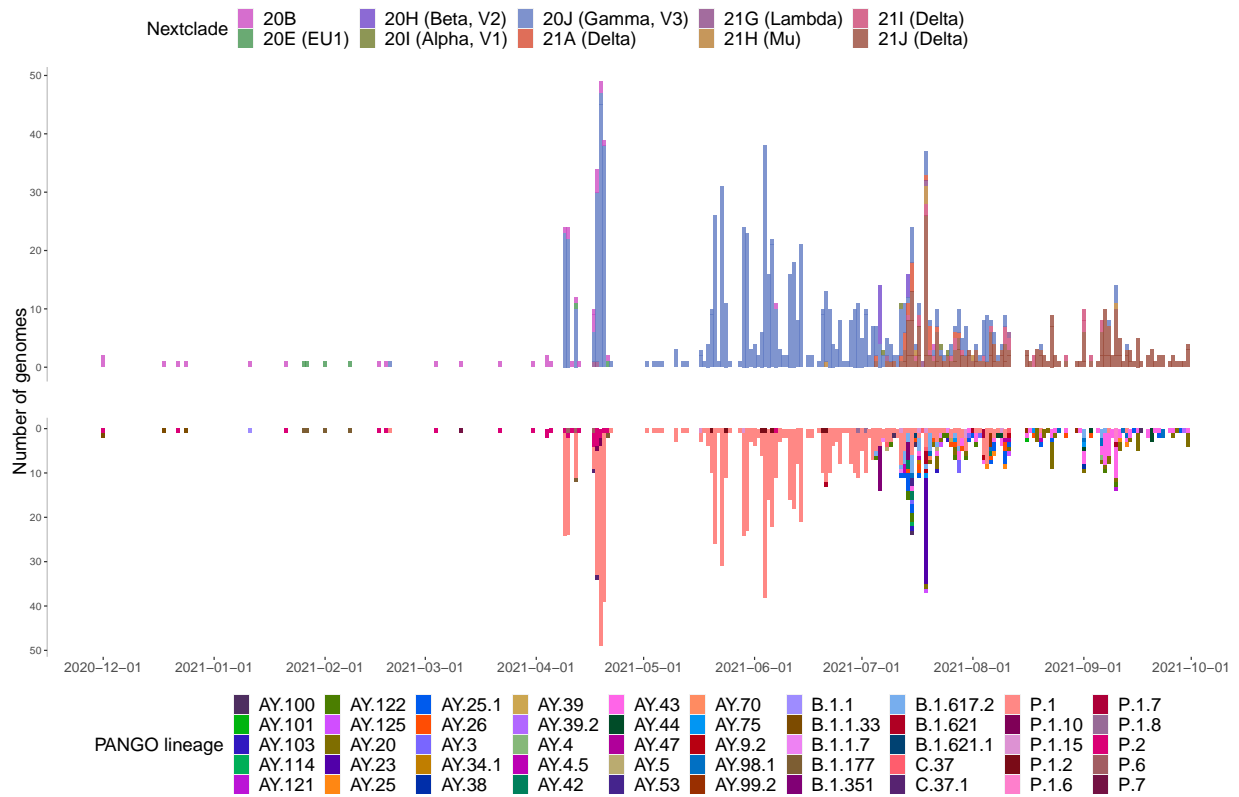


```
library(grid)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
## combine

sup2 <- annotate_figure(sup, left = textGrob("Number of genomes", rot = 90, vjust = 1,
  gp = gpar(cex = 1.3)))
sup2
```

```
png("Supplementary_3.png", res = 600, height = 25, width = 40, units = "cm")
sup2
dev.off()
```

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