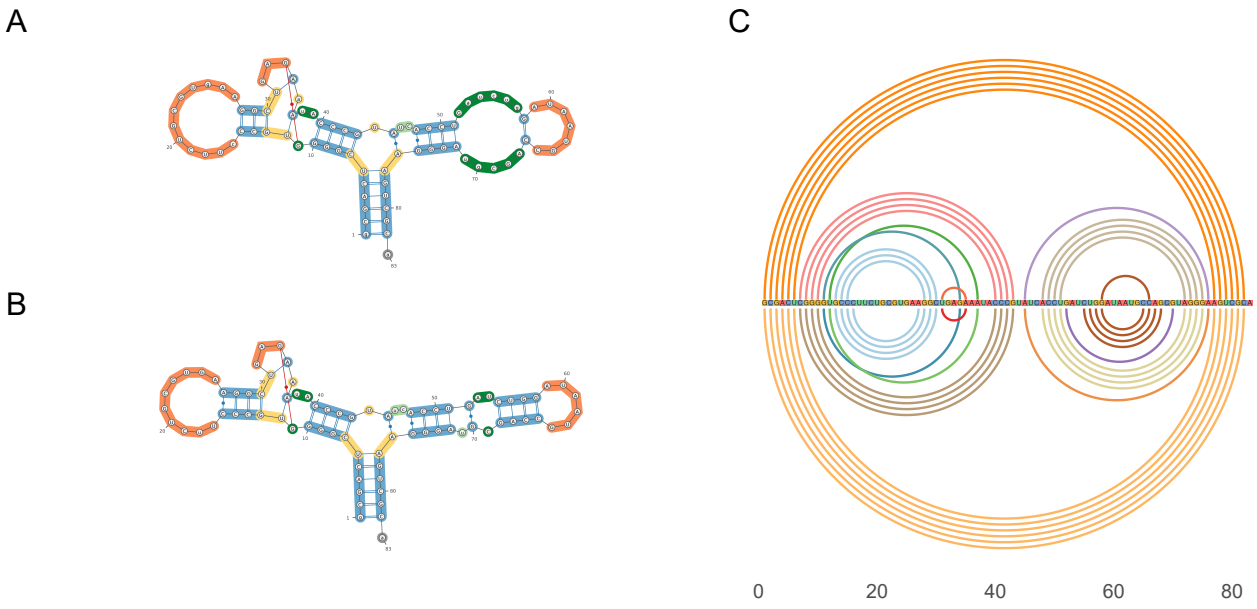
Supplemental File of

ggmsa: a visual exploration tool for multiple sequence alignment and associated data

```
library(ggmsa)
library(ggplot2)
library(magick)
library(ggplotify)
library(aplot)
library(RColorBrewer)
library(Biostrings)
library(treeio)
library(ggtree)
```

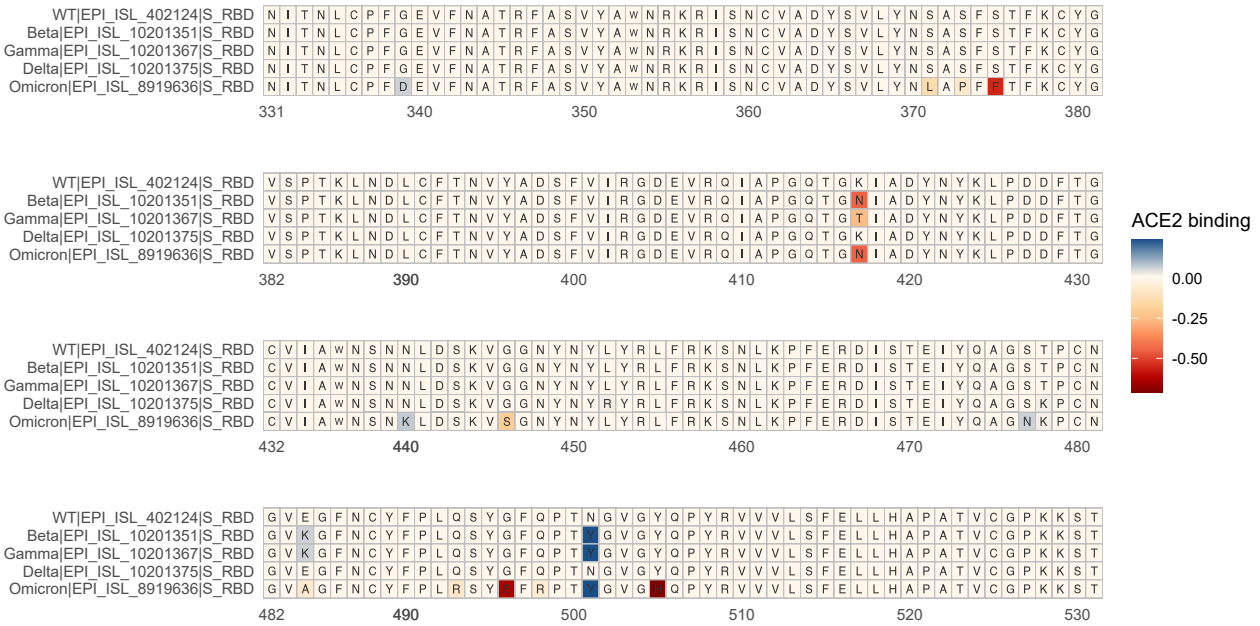
Reviewer1: Require1

```
tpp_seq <- "data/tpp_riboswitch.fasta"
arc_4NYG <- "data/riboswitch_thiamine.txt"
arc_4NYD <- "data/riboswitch_hypoxanthine.txt"
thiamine <- readSSfile(arc_4NYG, type = "Vienna" )
hypoxanthine <- readSSfile(arc_4NYD, type = "Vienna")

p <- ggmsa(tpp_seq,
           color = "Chemistry_NT",
           seq_name = F,
           show.legend = F,
           border = NA) +
    geom_helix(helix_data = list(known = hypoxanthine,
                                 predicted = thiamine)) +
    theme(axis.text.y = element_blank())

p1 <- image_read_pdf(path = "data/bpRNA_PDB_590_ColorCodedStructures_4NYG.pdf",
                     density = 300)
p2 <- image_read_pdf(path = "data/bpRNA_PDB_589_ColorCodedStructures_4NYD.pdf",
                     density = 300)

q1 <- as.ggplot(p1)
q2 <- as.ggplot(p2)

p_loop <- plot_list(gglist = list(q1, q2), ncol = 1)
pp <- plot_list(gglist = list(p_loop, p), ncol = 2, tag_levels = "A")
pp
```

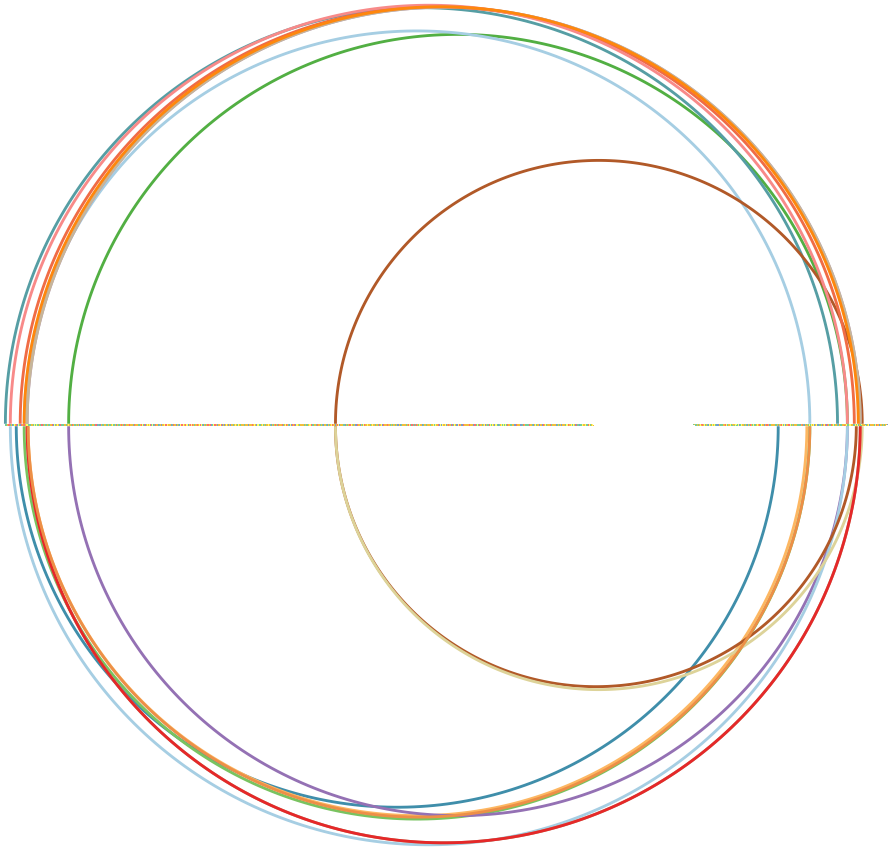


Reviewer1: Require2

```
colRD <- colorRampPalette(rev(brewer.pal(n = 9, name = "OrRd")))
colBU <- colorRampPalette(colors = rev(c("#185089","#FFF7EC")))

data <- "data/s_RBD.fasta"
mds <- read.csv("data/MDS.csv")
del <- c("expr_lib1", "expr_lib2",
         "expr_avg","bind_lib1",
         "bind_lib2")

mds <- mds[,!colnames(mds) %in% del]

tidymsa <- tidy_msa(data)
tidymsa <- assign_mds(tidymsa, mds)
tidymsa$position <- tidymsa$position + 330

p <- ggplot() +
    geom_msa(data = tidymsa,
             char_width = 0.5,
             mds = TRUE,
             seq_name = TRUE,
             show.legend = TRUE) +
    theme_msa() +
    scale_fill_gradientn(name = "ACE2 binding",
                         colors = c(colRD(75),colBU(25))) +
    facet_msa(50)
p
```

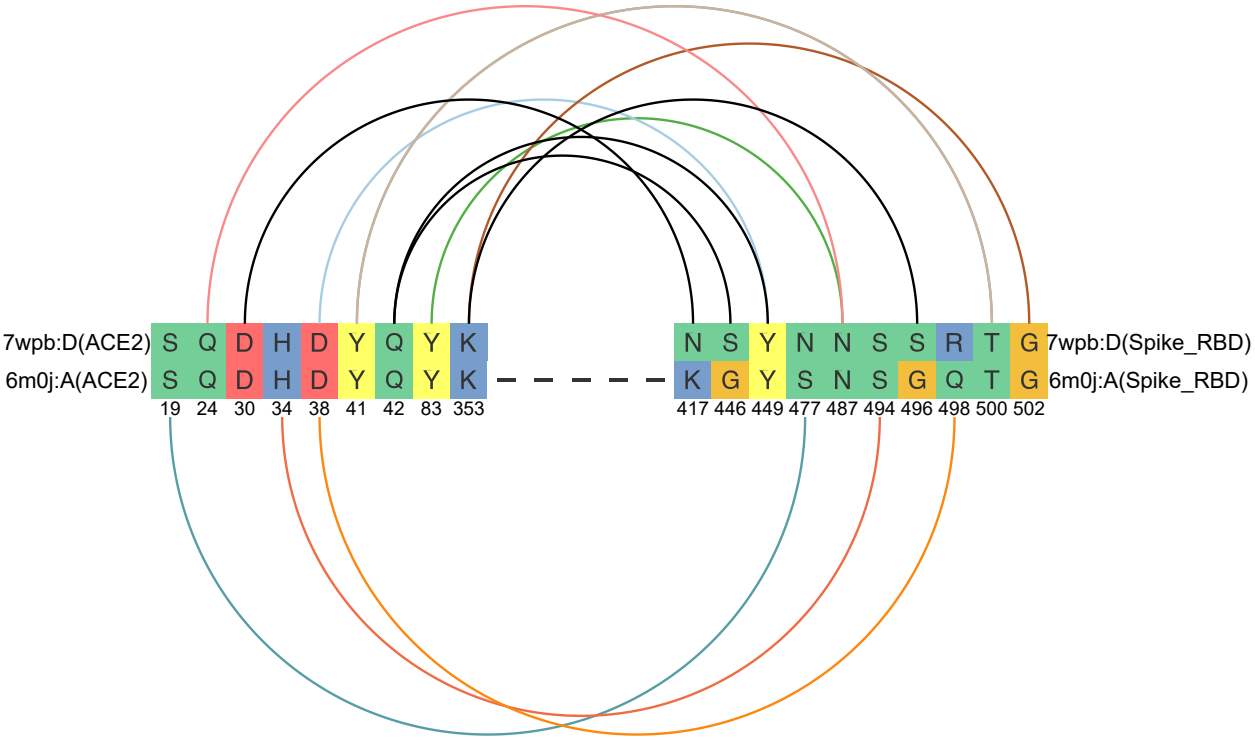



```
#read sequences
x <- readAAStringSet("data/ACE2.fasta")
y <- readAAStringSet("data/Spike_RBD.fasta")

#read protein-protein position
inter1 <- read.csv("data/6m0j_inter.csv")
inter2 <- read.csv("data/7wpb_inter.csv")

#tidy data
t1 <- tidy_msa(x, start = 19)
t2 <- tidy_msa(y, start = 331)
t_merge <- merge_seq(t1, gap = 100, t2)

h1 <- tidy_hdata(100, inter1, t1, t2)
h2 <- tidy_hdata(100, inter2, t1, t2)

#protein-protein interactive plot
p_interactive <- ggplot() +
    geom_msa(data = t_merge,border = NA, font = NULL, seq_name = FALSE) +
    theme_msa() + theme(axis.text = element_blank()) +
    geom_helix(helix_data = list(known = h1, predicted = h2))
p_interactive
```



```
#simplified p-p interactive  plot
ACE2 <- t_merge[t_merge$position %in% c(h1$i,h2$i),]
spike <- t_merge[t_merge$position %in% c(h1$j,h2$j),]
t1_reset <- reset_pos(ACE2)
t2_reset <- reset_pos(spike)
simplify_merge <- merge_seq(previous_seq = t1_reset,
                            subsequent_seq = t2_reset,
                            gap = 5,
                            adjust_name = FALSE)

sim_h1 <- simplify_hdata(hdata = h1, sim_msa = simplify_merge)
sim_h2 <- simplify_hdata(hdata = h2, sim_msa = simplify_merge)

##break and label
b <- simplify_merge[simplify_merge$character != "-", "position"] %>% unique()
l <-  c(inter1$Res.no.1,inter1$Res.no..2,
        inter2$Res.no.1,inter2$Res.no..2) %>% unique

p_sim <- ggplot() +
    geom_msa(data = simplify_merge,border = NA, char_width = 0.5, seq_name = F) +
    ggmsa:::theme_msa() +
    geom_helix(helix_data = list(known = sim_h1,
                                 predicted = sim_h2),
               overlap = T) +
    geom_text(mapping = aes(x = b,
                            y = 0.25,
                            label = l[order(l)]),
```

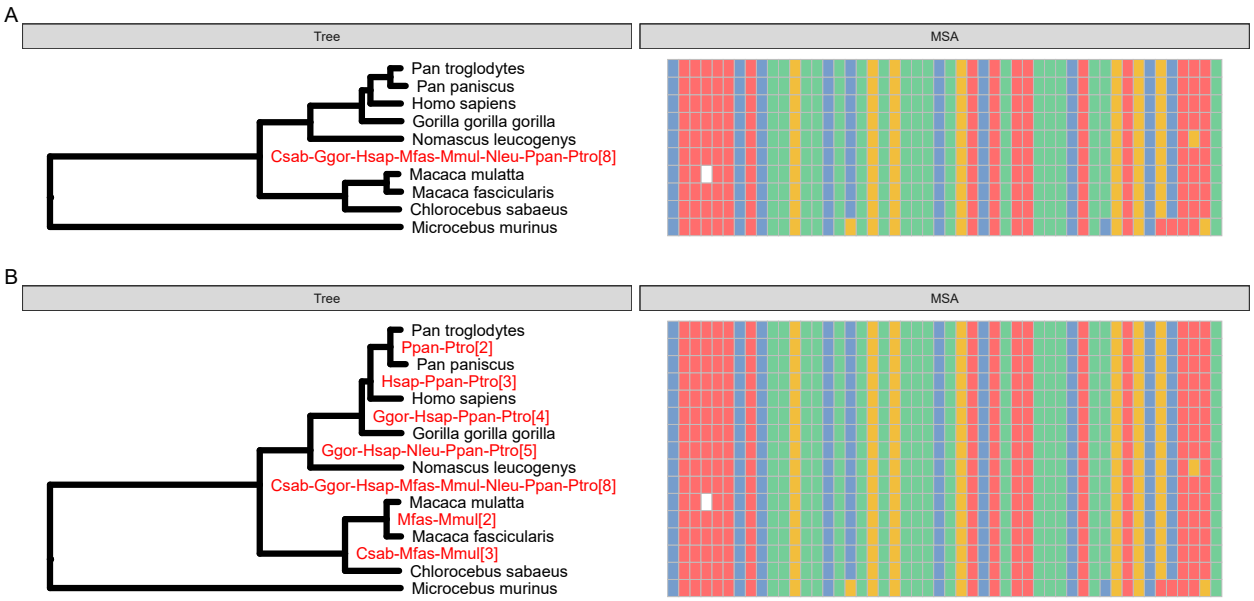
```
              size = 3) +
    theme(axis.text.x = element_blank(),
          axis.text.y = element_blank()) +
    scale_y_discrete(labels = c("6m0j:A(ACE2)", "7wpb:D(ACE2)")) +
    geom_text(aes(x = c(27.2,27.2),
                  y = c(1,2),
                  label = c("6m0j:A(Spike_RBD)",
                            "7wpb:D(Spike_RBD)")),
              size = 3.5) +
    geom_text(aes(x = c(-1.5,-1.5),
                  y = c(1,2),
                  label = c("6m0j:A(ACE2)",
                            "7wpb:D(ACE2)")),
              size = 3.5)  + xlim(-2,29)
p_sim
```



Reviewer2: Require1

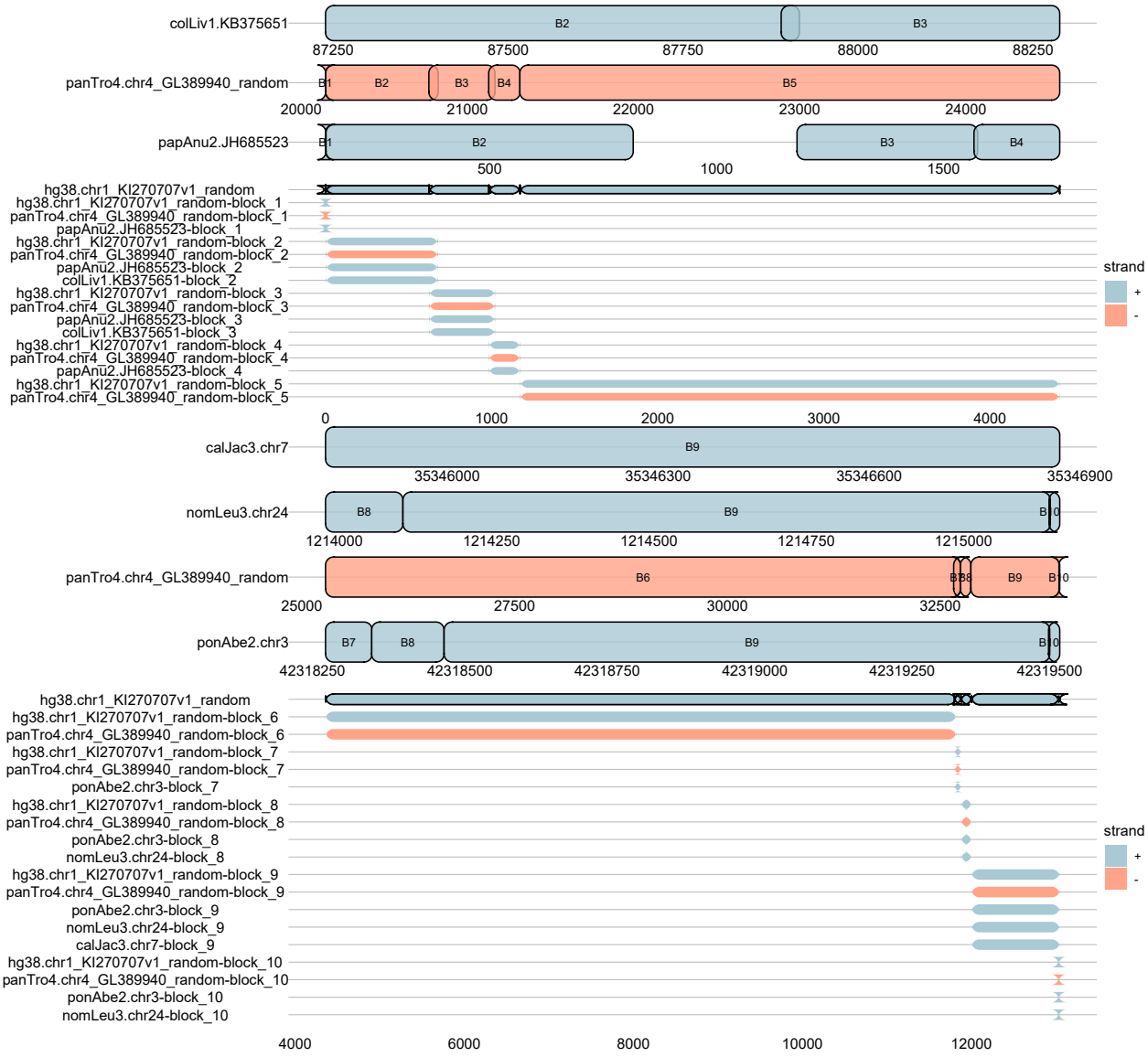
```
x <- read.phyloxml("data/msa_phyloxml.xml")
p <- ggtree(x, size = 2) + xlim_tree(0.12)
tidymsa <- extract_seq(p)

p1 <- treeMSA_plot(p,
                   tidymsa,
                   ancestral_node = 11,
                   sub = FALSE,
                   color = "Chemistry_NT")

p2 <- treeMSA_plot(p,
```

```
                    tidymsa,
                    ancestral_node = 11,
                    sub = TRUE,
                    color = "Chemistry_NT")

pp <- plot_list(gglist = list(p1,p2),
                nrow = 2,
                tag_levels = "A",
                heights = c(0.4,0.6))
pp
```



Reviewer2: Require2

```
maf <- "data/chr1_KI270707v1_random.txt.maf"
ref = "hg38.chr1_KI270707v1_random"
seq_df <- read_maf(maf)
tidy_df <- tidy_maf_df(seq_df, ref = ref)

ggmaf(data = tidy_df,
      ref = ref,
      block_start = 1,
      block_end = 10,
      facet_field = 5,
      facet_heights = c(0.4,0.6))
```


colLiv1.KB375651
B2
B3
87250
87500
87750
88000
88250
panTro4.chr4_GL389940_random
B1
B2
B3
B4
B5
20000
21000
22000
23000
24000
papAnu2.JH685523
B1
B2
B3
B4
500
1000
1500
hg38.chr1_KI270707v1_random
hg38.chr1_KI270707v1_random-block_1
panTro4.chr4_GL389940_random-block_1
papAnu2.JH685523-block_1
hg38.chr1_KI270707v1_random-block_2
panTro4.chr4_GL389940_random-block_2
papAnu2.JH685523-block_2
colLiv1.KB375651-block_2
hg38.chr1_KI270707v1_random-block_3
panTro4.chr4_GL389940_random-block_3
papAnu2.JH685523-block_3
colLiv1.KB375651-block_3
hg38.chr1_KI270707v1_random-block_4
panTro4.chr4_GL389940_random-block_4
papAnu2.JH685523-block_4
hg38.chr1_KI270707v1_random-block_5
panTro4.chr4_GL389940_random-block_5
0
1000
2000
3000
4000
strand
+
-
calJac3.chr7
B9
35346000
35346300
35346600
35346900
nomLeu3.chr24
B8
B9
1214000
1214250
1214500
1214750
1215000
panTro4.chr4_GL389940_random
B6
B9
25000
27500
30000
32500
ponAbe2.chr3
B7
B8
B9
42318250
42318500
42318750
42319000
42319250
42319500
hg38.chr1_KI270707v1_random
hg38.chr1_KI270707v1_random-block_6
panTro4.chr4_GL389940_random-block_6
hg38.chr1_KI270707v1_random-block_7
panTro4.chr4_GL389940_random-block_7
ponAbe2.chr3-block_7
hg38.chr1_KI270707v1_random-block_8
panTro4.chr4_GL389940_random-block_8
ponAbe2.chr3-block_8
nomLeu3.chr24-block_8
hg38.chr1_KI270707v1_random-block_9
panTro4.chr4_GL389940_random-block_9
ponAbe2.chr3-block_9
nomLeu3.chr24-block_9
calJac3.chr7-block_9
hg38.chr1_KI270707v1_random-block_10
panTro4.chr4_GL389940_random-block_10
ponAbe2.chr3-block_10
nomLeu3.chr24-block_10
4000
6000
8000
10000
12000
strand
+
-