Supplemental File of

Phylogenetic placement visualization with treeio-ggtree method

0.1 Favorable placement filtration after parsing jplace file

```r
# Required packages
library(treeio)
library(ggtree)
library(jsonlite)
library(dplyr)
library(ggtreeExtra)
library(RColorBrewer)
library(colorspace)
library(ggplot2)
library(aplot)

# Read jplace
jtree <- fromJSON("../exampledata/Holomycota/HolomycotaV4_alignedtrim.jplace")
phylo <- treeio:::jplace_treetext_to_phylo(jtree$tree)
placements <- treeio:::extract.placement(jtree,phylo)

# placement filtering
filtered_placements <- group_by(placements, .data$name) %>%
    filter(.data$like_weight_ratio == max(.data$like_weight_ratio))


p1 <- ggplot(data = placements,aes(x=like_weight_ratio)) +
    geom_histogram(data=filtered_placements,aes(x=like_weight_ratio,..count..),fill="skyblue",  color="white",s
    geom_histogram(data=placements,aes(x=like_weight_ratio,..count..),fill="red",  color="black",size=0.1,binwi
    labs(y="Number of Placements",x="LWR") +
    scale_x_continuous(breaks = c(seq(0,1,0.1))) +
    theme_bw()

tree1 <- as.treedata(phylo)
td <- as_tibble(tree1)
tree1@data <- group_by(filtered_placements, .data$node) %>% summarize(nplace=n()) %>%
    full_join(td, by='node') %>%
    mutate(nplace = ifelse(is.na(.data$nplace), 0, .data$nplace))

p2 <- ggtree(tree1,layout = "circular",branch.length = "none",aes(color=nplace,size=nplace)) +
    scale_color_continuous_sequential(palette = "Sunset",limits =c(1, 55)) +
    scale_size_continuous(range=c(0.1, 2),limits =c(0, 55))

V4_group <- read.csv(file = "../exampledata/Holomycota/V4_group.csv")
getPalette <- colorRampPalette(brewer.pal(9, "Set1"))
cols <- getPalette(21)
levels <- rev(unique(V4_group$group))

p3 <- p2%<+% V4_group +
    geom_fruit(
        geom = geom_tile,
        mapping = aes(fill = factor(group,levels = levels)),
        width = 2,
        offset = 0.1
    ) +
    scale_fill_manual(
        name = "",
        values = cols,
```
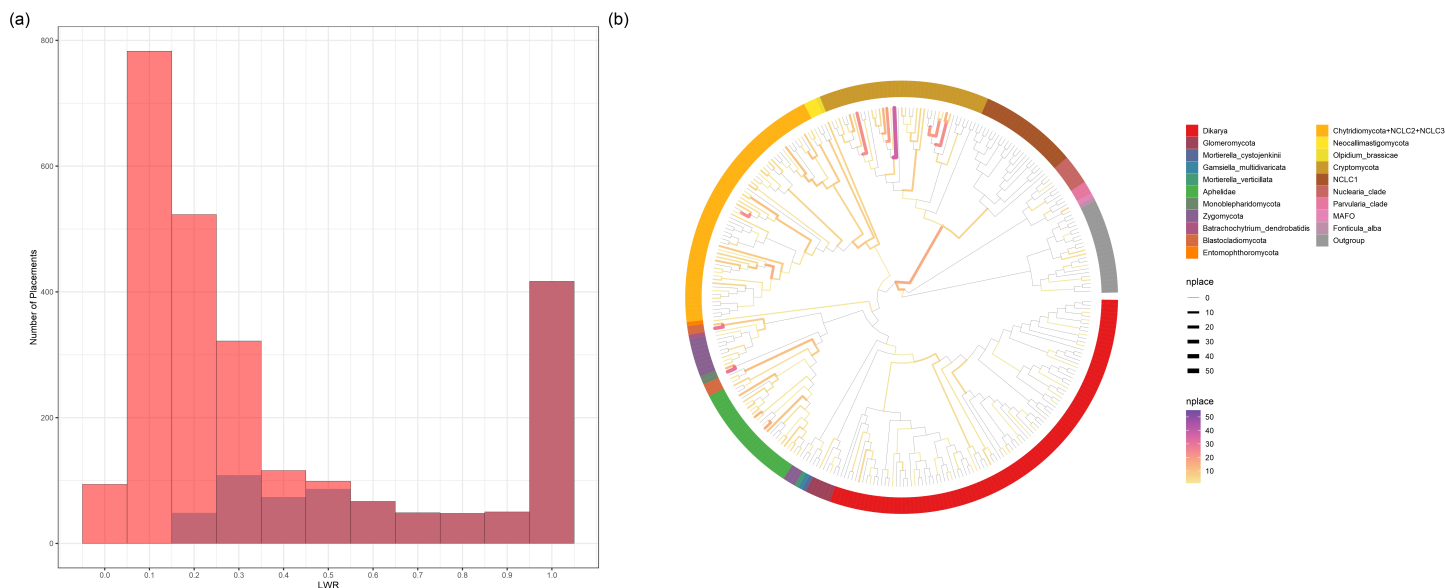
```
        guide = guide_legend(
            keywidth = 1,
            keyheight = 1,
            order = 3
        ),
        na.translate = FALSE
    )


plot_list(p1, p3,tag_levels = list(c('(a)', '(b)','(c)','(d)'), '1'), guides = 'collect',tag_size = 20)
```


0.2 Utility to explore the placement uncertainty

```
# Required packages
library(ggplot2)
library(treeio)
library(ggtree)
library(jsonlite)
library(dplyr)
library(colorspace)
library(ggtreeExtra)
library(aplot)
library(patchwork)

# Read jplace
jtree <- fromJSON("../exampledata/subtree/pplacer_Amt_subtree.jplace")
phylo <- treeio:::jplace_treetext_to_phylo(jtree$tree)
placements <- treeio:::extract.placement(jtree,phylo)
seq_name <- "saltern1"
sgplacement <- group_by(placements, .data$name) %>%
    filter(.data$name == seq_name)

tree1 <- as.treedata(phylo)
tree2 <- left_join(tree1,sgplacement,by="node")
p1 <- ggtree(tree2,layout = "circular",aes(color=like_weight_ratio)) +
    scale_color_continuous_sequential(palette = "Sunset",limits =c(0, 1)) +
    ggtitle(seq_name)
```

```
Amt_tiplabel <- read.csv(file = "../exampledata/subtree/Amt_tiplabel.csv",
                         header = TRUE)
head(Amt_tiplabel)
```

```
##   node      tiplabel              Group
## 1    1 Asterochloris Trebouxiophyceae
## 2    2 Asterochloris Trebouxiophyceae
## 3    3      Coccomyxa Trebouxiophyceae
## 4    4  Stichococcus Trebouxiophyceae
## 5    5      Galdieria      Cyanidiales
## 6    6      Galdieria      Cyanidiales
```
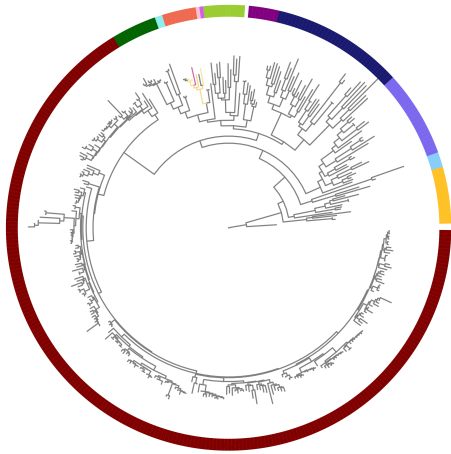
```
taxon_groups <- c("Trebouxiophyceae","Cyanidiales","Non-Cyanidiales",
                  "Chlamydomonadales","Tetraselmis ssp.","Pyramimonas ssp.",
                  "Nephroselmis pyriformis CCMP717","Bathycoccus ssp.",
                  "Ostreococcus","Mantoniella ssp.","Micromonas ssp.",
                  "Embryophyta")
Amt_tiplabel$Group <- factor(Amt_tiplabel$Group, levels=taxon_groups)
cols <- c("#FFC125","#87CEFA","#7B68EE","#191970","#800080",
          "#9ACD32","#D15FEE","#FFC0CB","#EE6A50","#8DEEEE",
          "#006400","#800000","#B0171F")


p2 <- p1%<+% Amt_tiplabel +
    geom_fruit(
        geom = geom_tile,
        mapping = aes(fill = Group),
        width = 0.2,
        offset = 0.01
    ) +
    scale_fill_manual(
        name = "",
        values = cols,
        guide = guide_legend(
            keywidth = 1,
            keyheight = 1,
            order = 3
        ),
        na.translate = FALSE
    )

# viewClade(p1,659) # viewClade
tree3 <- tree_subset(tree2, node=657, levels_back=0)
p3 <- ggtree(tree3,layout = "rect",aes(color=like_weight_ratio),size=1) +
    scale_color_continuous_sequential(palette = "Sunset",limits =c(0, 1)) +
    geom_tiplab(size=4,color="black",offset = 0.01) + xlim(0,2)



plot_list(p2, p3, tag_levels = list(c('(a)', '(b)'), '1'), guides = 'collect',
          nrow = 1,tag_size = 20)
```

(a) saltern1

(b)

CAMPEP_0179605098_Ostreococcus_mediterraneus_Strain_clade-D-RCC1621
CAMPEP_0179717148_Ostreococcus_mediterraneus_Strain_clade-D-RCC1107
CAMPEP_0179687388_Ostreococcus_mediterraneus_Strain_clade-D-RCC2573
CAMPEP_0179700822_Ostreococcus_mediterraneus_Strain_clade-D-RCC2593
UniRef100_H8ZJB2_Ostreococcus_tauri_virus_RT-2011
OstRCC809_2
UniRef100_UPI0000E4C67B_Ostreococcus_tauri
UniRef100_A4SAD9_Ostreococcus_lucimarinus_strain_CCE9901
CAMPEP_0179639450_Ostreococcus_prasinos_Strain_BCC99000
UniRef100_K8FCX2_Bathycoccus_prasinos

Trebouxiophyceae
Cyanidiales
Non-Cyanidiales
Chlamydomonadales
Tetraselmis ssp.
Pyramimonas ssp.
Nephroselmis pyriformis CCMP717
Bathycoccus ssp.
Ostreococcus
Mantoniella ssp.
Micromonas ssp.
Embryophyta

like_weight_ratio
1.00
0.75
0.50
0.25
0.00

0.3   Merge placement information to the placement tree

```r
# Required packages

library(treeio)
library(ggtree)
library(ggplot2)
library(ggtreeExtra)
library(colorspace)
library(dplyr)
library(patchwork)

tree1 <- read.jplace(file = "../exampledata/Mitsi/rsbl20190182supp2.jplace")
tree2 <- read.tree(file = "../exampledata/Mitsi/rsbl20190182supp7.tre")
tree2 <- as.treedata(tree2)

p1 <- ggtree(tree1,layout = "circular",aes(color=nplace,size=nplace)) +
    scale_color_viridis_c(option="viridis",limit=c(0.1,120)) +
    scale_size_continuous(range=c(0, 2))


# Extract placement informations
pla <- get.placements(tree1)

# merge placement data onto best-hit tree
#pla$label <- unlist(lapply(pla$name,function(x){paste("QUERY__",x,sep = "_")}))
pla$label <- pla$name
pla2 <- select(pla,c("name","label", "likelihood","like_weight_ratio"))
tree3 <- as_tibble(tree2)
tree4 <- left_join(tree3,pla2,by="label",)


tree4$group <- gsub("_1","",tree4$label)
tree4$group <- gsub("_2","",tree4$group)
tree4$group <- lapply(tree4$group,function(x){unlist(strsplit(x,"_"))[c(2)]})
tree4$group <- gsub("Rhabditophora","Bothrioplana_semperi",tree4$group)
s_group <- c("Catenulida","Polycladida","Macrostomorpha","Prorhynchidae",
            "Proseriata","Rhabdocoela","Bothrioplana_semperi","Fecampiidae",
            "Tricladida","Prolecithophora","Monogenea","Trematoda","Cestoda")
```

```r
otu_group <- unlist(list(tree4[grep("^OTU.",tree4$group),"group"]))
all_group <- c(s_group,otu_group)
tree4[!(tree4$group %in% all_group),"group"] <- "Outgroup"
tree4[seq(193,201),"group"] <- "Prorhynchidae"
tree4[seq(249,252),"group"] <- "Fecampiidae"
tree4[seq(169,171),"group"] <- "Gnosonesimidae"
tree4[seq(119,121),"group"] <- "Clade 1"
tree4[seq(174,192),"group"] <- "Clade 2"
s_group2 <- c("Trematoda", "Cestoda","Monogenea",
              "Prolecithophora", "Tricladida","Fecampiidae",
              "Bothrioplana_semperi","Rhabdocoela","Proseriata","Prorhynchidae",
              "Clade 2","Gnosonesimidae","Macrostomorpha","Polycladida",
              "Clade 1","Catenulida","Outgroup")
tree4[is.na(tree4$label),"group"] <- NA
tree4$group <- factor(tree4$group, levels=s_group2)
tree4[is.na(tree4$like_weight_ratio),"like_weight_ratio"] <- 0

tree5 <- as.treedata(tree4)

# tree5 <- root(tree5,node=572)
p2<- ggtree(tree5,
            layout = "circular",
            # branch.length = "none",

) + geom_tree(
    aes(
        color=like_weight_ratio,
        size=like_weight_ratio
    )
) +
    # geom_tiplab(size=2,offset=0.02,align=TRUE) +
    scale_color_continuous_sequential(palette = "Sunset", limits=c(0.000001, 0.2)) +
    scale_size_continuous(range=c(0, 1.5)) +
    geom_tiplab2(data=td_filter(like_weight_ratio > 0),align=TRUE,size=3,offset=0.1)

p3 <- scaleClade(p2,
         node=c(MRCA(tree5,"SA_OTU14871","SA_OTU158753")),
         scale=5)
p4 <- scaleClade(p3,
             node=c(MRCA(tree5,"SA_OTU22294","SA_OTU9482")),
             scale=5)




# Annotate Group Labels
#ggtree(tree5) + geom_text2(aes(x=branch,label=node))

p5 <- p4 +
    geom_fruit(geom = geom_tile,
             mapping = aes(fill = group),
             width = 0.1,
             offset = 0.01)
p5
```