SF8

NSMR

```r
library(readr)
library(ggplot2)
library(gridExtra)
library(grid)
library(gdata)
```

```
## gdata: read.xls support for 'XLS' (Excel 97-2004) files ENABLED.

##

## gdata: read.xls support for 'XLSX' (Excel 2007+) files ENABLED.

##
## Attaching package: 'gdata'

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

## The following object is masked from 'package:stats':
##
##     nobs

## The following object is masked from 'package:utils':
##
##     object.size

## The following object is masked from 'package:base':
##
##     startsWith
```

```r
library(ggpubr)
library(cowplot)
```

```
##
## Attaching package: 'cowplot'

## The following object is masked from 'package:ggpubr':
##
##     get_legend
```

```r
library(RColorBrewer)
library(pheatmap)
library(tidyr)
```

```r
df <- readr::read_csv('key_nodes.tidydf.csv')
```

```
## Parsed with column specification:
## cols(
##   node = col_character(),
```

```
##   taxon = col_character(),
##   species = col_character(),
##   random = col_character(),
##   block_id = col_double(),
##   iteration = col_double(),
##   density = col_double(),
##   acc_ls = col_character(),
##   all_acc_ls = col_character(),
##   total_density = col_double(),
##   total_genome_length = col_double(),
##   density_ratio = col_double(),
##   multi_sp = col_double(),
##   para = col_character(),
##   mean_dist_pair = col_double(),
##   mean_dist_pair_norm = col_double(),
##   median_dist_pair = col_double(),
##   median_dist_pair_norm = col_double()
## )
```

Here we define a function to make boxplots by taxon

```r
map_signif_level <- c(`****` = 1e-04, `***` = 0.001,  `**` = 0.01, `*` = 0.05, ns = 1)

make_plot <- function(tbl,
                      key = "observed",
                      comparisons = list(c("observed", "random")),
                      bracket_y = NULL,
                      ylims = c(0 , 0.0001)) {

  if(is.null(bracket_y)) {
    h = ylims[2] - ylims[1]
    bracket_y = c(.9,.825,.75)*h + ylims[1]
  }

  size.summary <- tbl %>% dplyr::filter(random == "observed") %>%  dplyr::group_by(taxon) %>%  dplyr::su
  ggplot(tbl, aes_string(x = 'random', y = 'median_dist_pair_norm', fill = 'random')) +
    geom_boxplot(outlier.shape = NA) +
    facet_grid(~ taxon) +
    theme_cowplot() +
    theme(axis.title.x = element_blank(), axis.text.x = element_blank()) +
    geom_signif(comparisons = comparisons,
                test = "wilcox.test", test.args = list(paired = FALSE, exact = FALSE), na.rm = TRUE,
                map_signif_level = map_signif_level,
                color="black", tip_length = 0.01, size = .5, textsize = 2,
                y_position = bracket_y,  data = NULL) +
    scale_y_continuous(name = "Median distance between consecutive genes (fraction of genome)", limits =
    theme(legend.title = element_blank(),
          plot.margin = unit(c(1,0,0,0), units='cm'),
          legend.position = 'bottom',
          legend.justification = 'center',
          strip.text = element_text(size = 6, angle = 90, margin = margin(5,0,5,0,'pt')),
          axis.ticks.x = element_blank(),
          axis.title.y = element_text(size = 7),
          axis.text = element_text(size = 6)) +
    geom_text(data=size.summary, aes(x=1,y= 0.000093,hjust = 0.5,label = label), size = 2, inherit.aes=
```

```
}

df$taxon <- factor(df$taxon, levels=c('Poriferan','Ctenophore','Placozoan','Cnidarian', 'Acoel','Ecdyso
df$node <- as.factor(df$node)
df$random <- as.factor(df$random)
df$species <- factor(df$species, levels = c('CAPOW', 'SALRO', 'AMPQU', 'SYCCI', 'MNELE', 'PLEBA', 'TRIA

df_para_Met <- df %>% dplyr::filter(para == 'para' & node == 'Metazoa')
df_para_Par <- df %>% dplyr::filter(para == 'para' & node == 'Parahoxozoa')
df_para_Pla <- df %>% dplyr::filter(para == 'para' & node == 'Planulozoa')
df_para_Bil <- df %>% dplyr::filter(para == 'para' & node == 'Bilateria')
df_para_Ver <- df %>% dplyr::filter(para == 'para' & node == 'Vertebrata')
df_para_Lop <- df %>% dplyr::filter(para == 'para'  & node == 'Lophotrochozoa')
df_not_para_Met <- df %>% dplyr::filter(para == 'not_para' & node == 'Metazoa')
df_not_para_Par <- df %>% dplyr::filter(para == 'not_para' & node == 'Parahoxozoa')
df_not_para_Pla <- df %>% dplyr::filter(para == 'not_para' & node == 'Planulozoa')
df_not_para_Bil <- df %>% dplyr::filter(para == 'not_para' & node == 'Bilateria')
df_not_para_Ver <- df %>% dplyr::filter(para == 'not_para' & node == 'Vertebrata')
df_not_para_Lop <- df %>% dplyr::filter(para == 'not_para' & node == 'Lophotrochozoa')
```

Every make_plot call for all the possibilities. Done so so that we can have ggpubr tests with facetting.

```
p1 <- make_plot(df_not_para_Met)

## `summarise()` ungrouping output (override with `.groups` argument)
p2 <- make_plot(df_para_Met)

## `summarise()` ungrouping output (override with `.groups` argument)
p3 <- make_plot(df_not_para_Par)

## `summarise()` ungrouping output (override with `.groups` argument)
p4 <- make_plot(df_para_Par)

## `summarise()` ungrouping output (override with `.groups` argument)
p5 <- make_plot(df_not_para_Pla)

## `summarise()` ungrouping output (override with `.groups` argument)
p6 <- make_plot(df_para_Pla)

## `summarise()` ungrouping output (override with `.groups` argument)
p7 <- make_plot(df_not_para_Bil)

## `summarise()` ungrouping output (override with `.groups` argument)
p8 <- make_plot(df_para_Bil)

## `summarise()` ungrouping output (override with `.groups` argument)
p9 <- make_plot(df_not_para_Ver)

## `summarise()` ungrouping output (override with `.groups` argument)
p10 <- make_plot(df_para_Ver)

## `summarise()` ungrouping output (override with `.groups` argument)
```
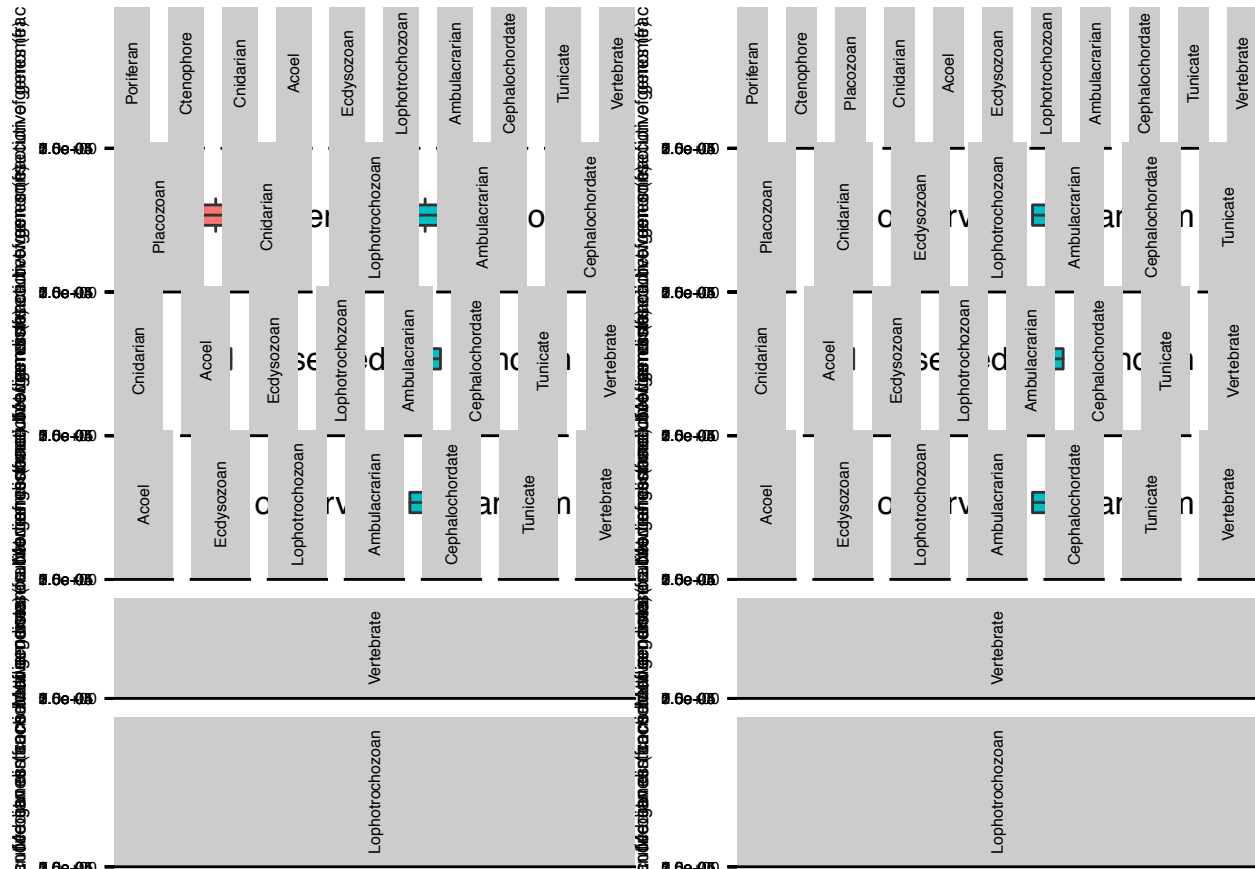
```r
p11 <- make_plot(df_not_para_Lop)
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
p12 <- make_plot(df_para_Lop)
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
gridplot <- gridExtra::grid.arrange(grobs = list(p1,p2,p3,p4,p5,p6,p7,p8,p9,p10,p11,p12), ncol = 2)
```

```
## Warning: Removed 73 rows containing non-finite values (stat_boxplot).
## Warning: Removed 322 rows containing non-finite values (stat_boxplot).
## Warning: Removed 62 rows containing non-finite values (stat_boxplot).
## Warning: Removed 423 rows containing non-finite values (stat_boxplot).
## Warning: Removed 680 rows containing non-finite values (stat_boxplot).
## Warning: Removed 834 rows containing non-finite values (stat_boxplot).
## Warning: Removed 639 rows containing non-finite values (stat_boxplot).
## Warning: Removed 272 rows containing non-finite values (stat_boxplot).
## Warning: Removed 34 rows containing non-finite values (stat_boxplot).
## Warning: Removed 207 rows containing non-finite values (stat_boxplot).
## Warning: Removed 107 rows containing non-finite values (stat_boxplot).
```

```
ggsave(plot = gridplot,
       filename = 'SF5_consecutivepairs_taxons_meds.pdf',
       unit = 'cm',
       width = 30,
       height = 80)
```