

Figure 4 - wnt regression lines

NSMR

```
library(readr)
library(ggplot2)
library(ggrepel)
library(ggpubr)
library(cowplot)

##
## Attaching package: 'cowplot'
## The following object is masked from 'package:ggpubr':
##
##   get_legend
library(patchwork)

##
## Attaching package: 'patchwork'
## The following object is masked from 'package:cowplot':
##
##   align_plots

Now we'll load up the data
df_wg_stats <- readr::read_csv("genome_stats.csv")

## Parsed with column specification:
## cols(
##   species = col_character(),
##   genome_length = col_double(),
##   gene_count = col_double()
## )

df_wnt <- readr::read_csv("wnt.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(),
##   total_density = col_double(),
##   density_ratio = col_double(),
##   multi_sp = col_character(),
```

```
## para = col_character()
## )

df_hox <- readr::read_csv("hox.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(),
##   total_density = col_double(),
##   density_ratio = col_double(),
##   multi_sp = col_character(),
##   para = col_character()
## )
```

Now we'll add some columns to wgd stats, drop the cols we don't need in wnt and hox df to cleanup

```
df_wg_stats$recip_wgd <- df_wg_stats$genome_length / df_wg_stats$gene_count

prep_my_df <- function(tbl){
  outdf <- tbl %>%
    dplyr::filter(random == 'observed') %>%
    dplyr::mutate(invertebrate = dplyr::case_when(
      taxon == 'Vertebrate' ~ 'Vertebrate',
      taxon != 'Vertebrate' ~ 'Invertebrate'),
      recip_block_density = 1 / density) %>%
    dplyr::select(c(taxon, species, density, total_density, invertebrate, recip_block_density)) %>%
    dplyr::left_join(df_wg_stats, by = 'species')
  outdf$taxon <- as.factor(outdf$taxon)
  return(outdf)
}

df_wnt <- prep_my_df(df_wnt)
df_hox <- prep_my_df(df_hox)
```

Now this is for getting linear fit, also preparing the palette

```
wg_lm <- lm(formula = recip_wgd ~ genome_length,
            data = df_wg_stats)
wg_lm2 <- lm(formula = recip_wgd / 2 ~ genome_length,
            data = df_wg_stats)

mypalette <- list('#1CA1FB', '#CC52AB', '#FA7850', '#AB1E3D', '#32B559', '#E61F00', '#106E82')
names(mypalette) <- c('Cnidarian', 'Tunicate', 'Ecdysozoan', 'Lophotrochozoan', 'Ambulacrarian', 'Cephalochordate')
```

This is to prepare for shading between the two ablines we provide

```
slope_wgd <- coef(wg_lm)[[2]]
slope_2_wgd <- coef(wg_lm2)[[2]]
intercept_wgd <- coef(wg_lm)[[1]]
```

```

intercept_2_wgd <- coef(wg_lm2)[[1]]

df_wg_stats$estimated_recip_wgd <- df_wg_stats$genome_length* slope_wgd + intercept_wgd
df_wg_stats$twice_estimated_recip_wgd <- df_wg_stats$genome_length* slope_2_wgd + intercept_2_wgd

plot with regression as boundaries
make_plot <- function(tbl){
  ggplot(tbl, aes_string(x = 'genome_length', y = 'recip_block_density', color = 'taxon')) +
    ggplot2::geom_ribbon(aes(x = genome_length,
                           ymin = estimated_recip_wgd,
                           ymax = twice_estimated_recip_wgd),
                        data = df_wg_stats,
                        inherit.aes = F,
                        fill = 'grey90',
                        color = 'grey80',
                        linetype = 2)+
    ggplot2::scale_color_manual(values = mypalette)+
    ggplot2::geom_smooth(method = 'lm', se = F, size = 0.5) +
    ggplot2::geom_point()+
    cowplot::theme_cowplot() +
    ggplot2::scale_x_continuous(name = 'Assembly size (bp)',
                               trans = 'log10') +
    ggplot2::scale_y_continuous(name = 'Reciprocal of gene density (bp/gene)',
                               trans = 'log10')+
    ggplot2::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.title.x = element_text(size = 7),
                   axis.title.y = element_text(size = 7),
                   axis.text = element_text(size = 6),
                   legend.text = element_text(size = 7))
}

wnt_p <- make_plot(df_wnt)
hox_p <- make_plot(df_hox)

prow <- cowplot::plot_grid(wnt_p + ggplot2::theme(legend.position="none"),
                           hox_p + ggplot2::theme(legend.position="none"),
                           align = 'vh',
                           labels = c("A", "B", "C"),
                           hjust = -1,
                           nrow = 1)

## `geom_smooth()`` using formula 'y ~ x'
## `geom_smooth()`` using formula 'y ~ x'
#legend_b <- get_legend(hox_p + theme(legend.position="bottom"))

p <- cowplot::plot_grid(prow, ncol = 1, rel_heights = c(1, .2))

ggsave(plot = p,
        filename = 'hox_wnt_regressions.pdf',

```

```
unit = 'cm',  
width = 15,  
height = 7)
```