SF5

NSMR

```r
library(readr)
library(ggplot2)
library(ggrepel)
library(ggpubr)
library(scales)
```

```
##
## Attaching package: 'scales'

## The following object is masked from 'package:readr':
##
##     col_factor
```

```r
library(cowplot)
```

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

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

Now we'll load up the data, change iteration/phylo_hypothesis to factor type data

```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()
## )
```

```
df_stats <- readr::read_csv("genome_stats.csv")

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

df <- dplyr::left_join(df, df_stats, by = 'species')
```

Plot reciprocal genome density as a funtion of assembly size

```
df$reciprocal_total_density <- 1 / df$total_density

p<- ggplot2::ggplot(df, aes_string(x= 'genome_length', y = 'reciprocal_total_density')) +
    ggplot2::geom_smooth(method = lm) +
    ggplot2::geom_point() +
    ggpubr::stat_cor(size = 2, label.y = 175000) +
    ggpubr::stat_regline_equation(size = 2, label.y = 200000) +
    ggplot2::scale_y_continuous(name = "Reciprocal of whole genome density") +
    ggplot2::scale_x_continuous(name = "Assembly size") +
    cowplot::theme_cowplot(font_size = 8)

ggsave(filename = 'SF2_Metazoa_assembly_size_total_density.pdf',
       plot = p,
       units = 'cm',
       width = 10,
       height = 6)
```

```
## `geom_smooth()` using formula 'y ~ x'
```