Figure 1 and supplements

NSMR

```
library(readr)
library(ggplot2)
library(ggrepel)
library(ggpubr)
library(pheatmap)
library(gridExtra)
library(RColorBrewer)
```

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

```
df <- readr::read_delim("node_counts.tsv", delim = "\t")
```

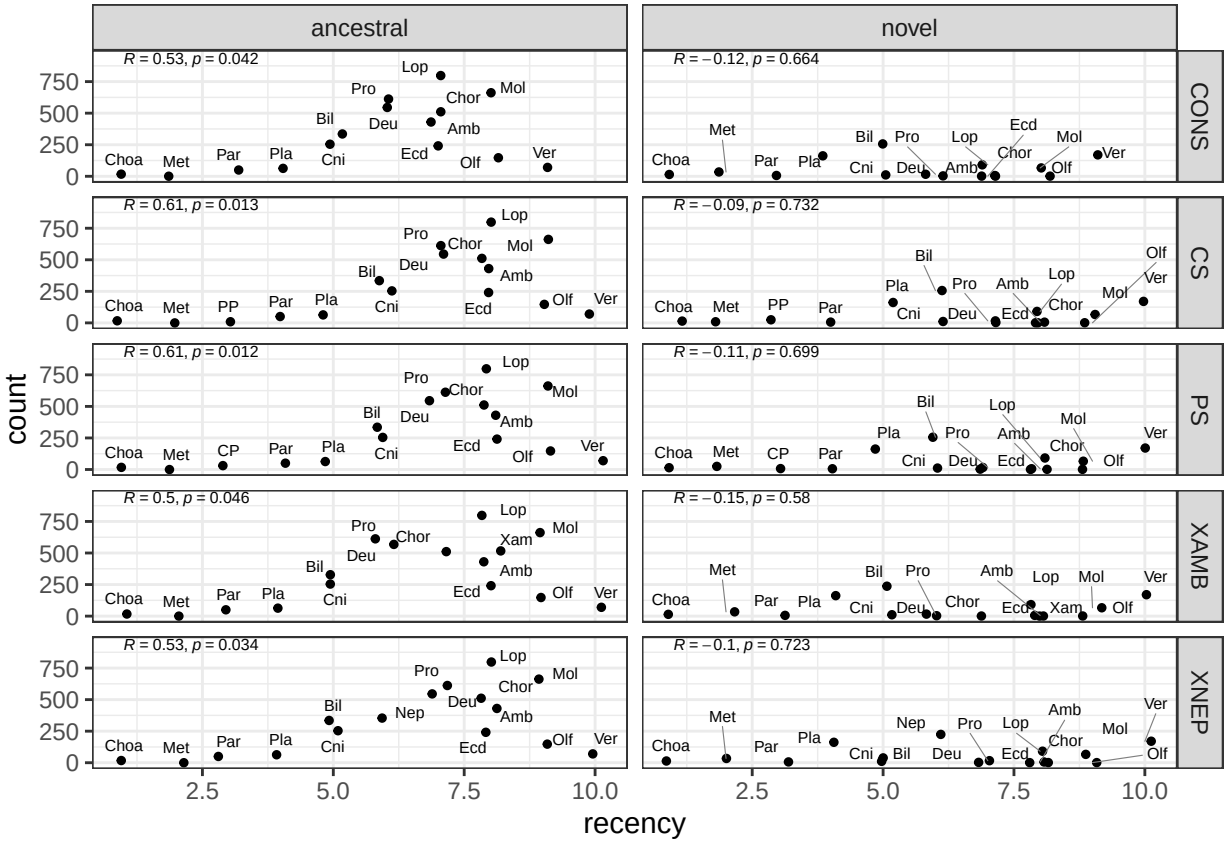
```
## Parsed with column specification:
## cols(
##   node_name = col_character(),
##   node_short = col_character(),
##   block_type = col_character(),
##   count = col_double(),
##   phylo_hypothesis = col_character(),
##   recency = col_double(),
##   observed = col_character(),
##   iteration = col_double()
## )
```

```
df$iteration <- as.factor(df$iteration)
df$block_type <- as.factor(df$block_type)
df$node_name <- as.character(df$node_name)
df$recency <- as.integer(df$recency)
df$phylo_hypothesis <- as.factor(df$phylo_hypothesis)

df_obs <- dplyr::filter(df, observed == 'obs')
df_rand <- dplyr::filter(df, observed == 'rand')
```

observed blocks counts as a function of recency

```
obs_scatter <- ggplot(df_obs, aes(x= recency, y = count, ymax = 950, label = node_short)) +
                geom_point(position = position_jitter(width = 0.2), size = 1) +
                ggrepel::geom_text_repel(size = 2,
                                         segment.size  = 0.2,
                                         segment.color = "grey50") +
                ggpubr::stat_cor(p.accuracy = 0.001,
                                 r.accuracy = 0.01,
                                 method = 'spearman',
                                 label.x = 1,
                                 label.y = 925,
                                 color = 'black',
```

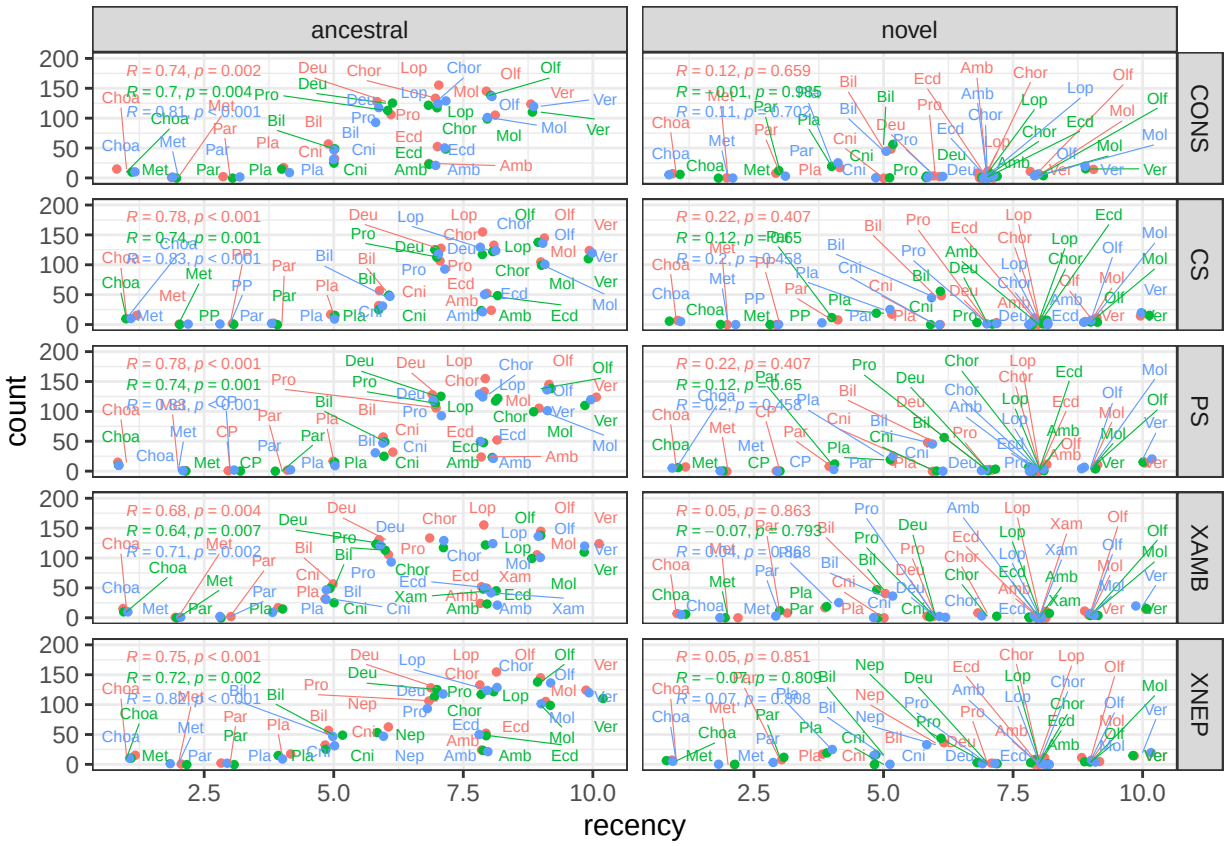
```
                                size = 2)


obs_scatter + facet_grid(phylo_hypothesis ~ block_type) + theme_bw()
```




```
ggsave(filename = 'SF1A_obs_block_counts_recency.pdf',
       units = 'cm',
       width = 9,
       height = 16)
```

random blocks counts as a function of recency (supplementary figure)

```
rand_scatter <- ggplot(df_rand, aes(x= recency, y = count, ymax = 200, color = iteration, label = node_s
                geom_point(position = position_jitter(width = 0.2), size = 1) +
                ggrepel::geom_text_repel(size = 2,
                                         segment.size  = 0.2) +
                guides(color = FALSE) + #remove legend, iteration nb is not important
                ggpubr::stat_cor(p.accuracy = 0.001,
                                 r.accuracy = 0.01,
                                 method = 'spearman',
                                 size = 2)


rand_scatter + facet_grid(phylo_hypothesis ~ block_type) + theme_bw()
```




```
ggsave(filename = 'SF1B_rand_block_counts_recency.pdf',
       units = 'cm',
       width = 21,
       height = 29.7)
```

retention matrix used for the heatmap in figure 1 We use the scatterplot density data used in figure 2

```
df <- readr::read_csv('raw_data_scatter.csv')
```

```
## Parsed with column specification:
## cols(
##   multi_sp = col_double(),
##   node = col_character(),
##   Vertebrate = col_double(),
##   Tunicate = col_double(),
##   Cephalochordate = col_double(),
##   Ambulacrarian = col_double(),
##   Lophotrochozoan = col_double(),
##   Ecdysozoan = col_double(),
##   Acoel = col_double(),
##   Cnidarian = col_double(),
##   Placozoan = col_double(),
##   Ctenophore = col_double(),
##   Poriferan = col_double()
## )
```

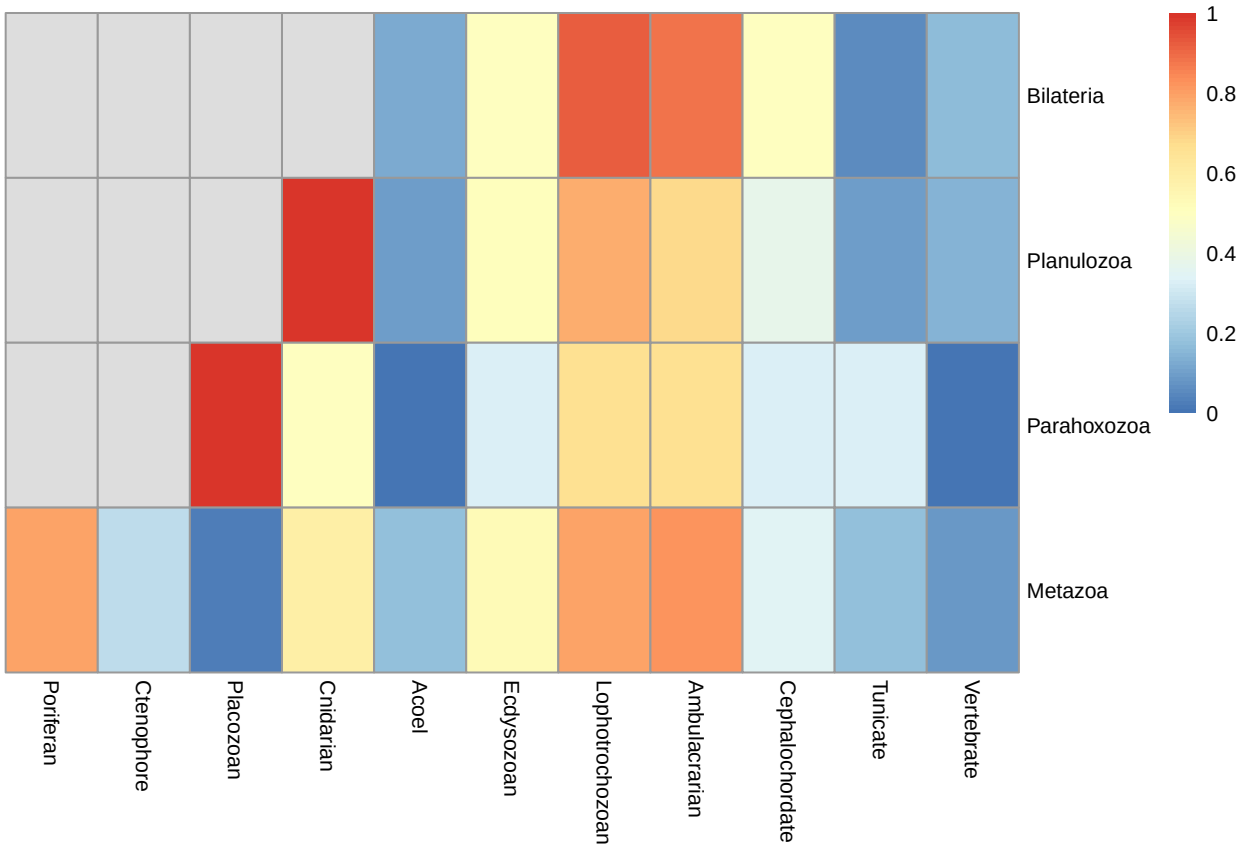
```
node_names <- c('Bilateria', 'Planulozoa','Parahoxozoa', 'Metazoa')
meta_names <- c("Poriferan", "Ctenophore", "Placozoan", "Cnidarian", "Acoel", "Ecdysozoan", "Lophotroch
```

```
ancestral_retention_matrix <- matrix(nrow = length(node_names), ncol = length(meta_names))
colnames(ancestral_retention_matrix) <- meta_names
rownames(ancestral_retention_matrix) <- node_names

for (mytaxon in meta_names){
  for (mynode in node_names){
    node_df <- df %>% dplyr::filter(node == mynode)
    df_retained_blocks <- node_df[mytaxon] %>% na.omit()
    retention_percent <- dplyr::tally(df_retained_blocks)$n / dplyr::tally(node_df)$n
    ancestral_retention_matrix[mynode, mytaxon] <- retention_percent
  }
}


ancestral_retention_matrix['Parahoxozoa', c('Poriferan', 'Ctenophore')] <- NA
ancestral_retention_matrix['Planulozoa', c('Poriferan', 'Ctenophore', 'Placozoan')] <- NA
ancestral_retention_matrix['Bilateria', c('Poriferan', 'Ctenophore', 'Placozoan', 'Cnidarian')] <- NA
```

Now for plotting the actual heatmap

retention heatmap that goes in figure 1

```
bk2 <- c(seq(0, 1,  length=100))

plot <-pheatmap::pheatmap(ancestral_retention_matrix,
                          cluster_rows = F,
                          cluster_cols = F,
                          breaks = bk2,
                          fontsize = 8)
```




```
ggsave(plot = plot,
       filename = 'fig1_heatmaps_retention.pdf',
       unit = 'cm',
       width = 15,
       height = 5)
```