# SpeciesOrthologBreakdown

## Purpose

To generate stacked bar graphs for the NHP species of interest showing how many of each human ortholog type there are for that species.

```r
library(dplyr)
library(stringr)
library(ggplot2)
library(reshape2)
library(purrr)
library(plyr)
library(AMR)
```

Upload file previously generated that contains the ortholog information for every annotated gene for the various species of interest.

```r
all_homologs <- get(load("2019-07-04homology_feature_NHP.Rdata"))
all_homologs_simple <- llply(all_homologs,
function(x) dplyr::select(x, ensembl_gene_id, hsapiens_homolog_orthology_type)) %>%
    llply(., function(x) distinct(x, ensembl_gene_id, .keep_all = TRUE)) %>%
  melt()
```

```
## Using ensembl_gene_id, hsapiens_homolog_orthology_type as id variables
## Using ensembl_gene_id, hsapiens_homolog_orthology_type as id variables
## Using ensembl_gene_id, hsapiens_homolog_orthology_type as id variables
## Using ensembl_gene_id, hsapiens_homolog_orthology_type as id variables
## Using ensembl_gene_id, hsapiens_homolog_orthology_type as id variables
## Using ensembl_gene_id, hsapiens_homolog_orthology_type as id variables
## Using ensembl_gene_id, hsapiens_homolog_orthology_type as id variables
## Using ensembl_gene_id, hsapiens_homolog_orthology_type as id variables
```

```r
##NHP genes without a human ortholog are currently just blank. Changing them to read
##"none" instead.
all_homologs_simple$hsapiens_homolog_orthology_type <-
  ifelse(all_homologs_simple$hsapiens_homolog_orthology_type == "", "none",
         all_homologs_simple$hsapiens_homolog_orthology_type)

homolog_freq <- all_homologs_simple %>% freq(hsapiens_homolog_orthology_type, L1)
homolog_freq$species <- str_extract(homolog_freq$item, "\\w+$")
homolog_freq$type <- str_extract(homolog_freq$item, "^\\w+\\s")

##Reordering the levels so they are in the order we want in the graph
homolog_freq$species <- factor(homolog_freq$species,
        levels = c("chimp", "bonobo", "gorilla","orang", "rhesus", "ptmac", "bab", "sqmonk"))

##Plotting summary of human ortholog types for each species
plotting_summary <- ggplot(homolog_freq, aes(x = species, y = count, fill = type)) +
  geom_col(color = "black") +
  xlab("Species") + ylab ("Number of ENSEMBL IDs") +
  scale_y_continuous(breaks = seq(0, 35000, 5000), limits = c(0, 35000)) +
  theme(panel.background = element_rect(fill = "white"),
        panel.grid.major = element_line(colour = "black")) +
```

```r
  theme(axis.text.y = element_text(size = rel(1)), axis.title.y = element_text(size = rel(1.2)),
        legend.position = "right", axis.title.x = element_text(size = rel(1.2)),
        axis.text.x = element_text(size = rel(1.2))) +
  theme(strip.background = element_rect(colour="black", fill="white", linetype="solid"),
                              strip.text = element_text(size = 10))
ggsave(filename = paste(Sys.Date(), "HumanOrthoTypeSummary_NHPs.pdf"),
       plot = plotting_summary, width = 12, height = 6, device = "pdf")
```

Session Info

```r
sessionInfo()
```

```
## R version 3.5.2 (2018-12-20)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Sierra 10.12.6
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] bindrcpp_0.2.2 AMR_0.7.0      plyr_1.8.4     purrr_0.2.5
## [5] reshape2_1.4.3 ggplot2_3.1.0  stringr_1.3.1  dplyr_0.7.8
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.0          pillar_1.3.1          compiler_3.5.2
##  [4] bindr_0.1.1         tools_3.5.2           digest_0.6.18
##  [7] evaluate_0.12       tibble_2.0.1          gtable_0.2.0
## [10] pkgconfig_2.0.2     rlang_0.3.1           microbenchmark_1.4-6
## [13] yaml_2.2.0          xfun_0.4              withr_2.1.2
## [16] knitr_1.21          hms_0.4.2             grid_3.5.2
## [19] tidyselect_0.2.5    glue_1.3.0            data.table_1.12.0
## [22] R6_2.3.0            rmarkdown_1.11        magrittr_1.5
## [25] backports_1.1.3     scales_1.0.0          htmltools_0.3.6
## [28] assertthat_0.2.0    colorspace_1.4-0      stringi_1.2.4
## [31] lazyeval_0.2.1      munsell_0.5.0         crayon_1.3.4
```