

# Genome matches

*O. Denas*

*February 6, 2017*

Overlaid histograms of the  $MS_{homo sapiens, x}$  ( $x$  is *Mus\_musculus* and *Danio Rerio*).

```
library(dplyr)
library(ggplot2)

load_dt <- function(fname){
  chrom_ = strsplit(strsplit(fname, "[.]")[[1]][5], "__")[[1]][1]
  t_species = strsplit(fname, "[.]")[[1]][1]
  s_species = strsplit(strsplit(fname, "[.]")[[1]][5], "__")[[1]][2]
  read.csv(sprintf('new_genome_data/%s', fname),
    sep = ' ', col.names = c('match_length', 'count')) %>%
    mutate(chrom = chrom_, fname = fname,
      t_species = t_species, s_species = s_species)
}

hs_dr <- list.files('./new_genome_data/',
  pattern = "Hom.*_Da.*hist")
hs_mm <- list.files('./new_genome_data/',
  pattern = "Hom.*_Mu.*hist")

ggplot(do.call(rbind, lapply(c(hs_dr, hs_mm), load_dt)),
  aes(match_length, count, fill=s_species)) +
  geom_bar(stat='identity') +
  facet_wrap(~chrom) + scale_y_log10() +
  labs(title='Homom Sapiens vs. Danio Rerio and Mus Musculus',
    subtitle='By Chromosome')
```

# Homom Sapiens vs. Danio Rerio and Mus Musculus

By Chromosome

