

Genome report Plots

2024-04-23

#Plots

```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v lubridate  1.9.3      v tibble    3.2.1
## v purrr      1.0.2      v tidyr     1.3.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
##
## The following object is masked from 'package:tidyr':
##
##      smiths
```

```
library(ggnewscale)
library(ggpubr)
```

```
tb <- read.csv("all_genomes_stats.csv")

theme_set(
  theme_bw() +
  theme(legend.position = "top")
)

# Extract genus from species names
tb$Genus <- sapply(str_split(tb$Species, "_"), `[`, 1)
tb$Genus
```

```
## [1] "Tetrapedia"      "Agapostemon"     "Andrena"          "Andrena"
## [5] "Andrena"         "Andrena"         "Anthidium"        "Apis"
## [9] "Apis"           "Apis"           "Apis"            "Apis"
## [13] "Augochlora"      "Augochlorella"   "Bombus"           "Bombus"
```

```
## [17] "Bombus"      "Bombus"      "Bombus"      "Bombus"
## [21] "Bombus"      "Bombus"      "Bombus"      "Bombus"
## [25] "Bombus"      "Bombus"      "Bombus"      "Bombus"
## [29] "Bombus"      "Bombus"      "Bombus"      "Bombus"
## [33] "Bombus"      "Bombus"      "Bombus"      "Bombus"
## [37] "Bombus"      "Bombus"      "Bombus"      "Bombus"
## [41] "Bombus"      "Bombus"      "Bombus"      "Bombus"
## [45] "Bombus"      "Bombus"      "Ceratina"     "Ceratina"
## [49] "Ceratina"     "Colletes"    "Ctenoplectra" "Dufourea"
## [53] "Eufriesea"    "Euglossa"    "Exoneura"     "Exoneurella"
## [57] "Frieseomelitta" "Habropoda"   "Halictus"     "Halictus"
## [61] "Halictus"     "Halictus"    "Heterotrigona" "Holcopasites"
## [65] "Lasioglossum" "Lasioglossum" "Lasioglossum" "Lasioglossum"
## [69] "Lasioglossum" "Lasioglossum" "Lasioglossum" "Lasioglossum"
## [73] "Lasioglossum" "Lasioglossum" "Lasioglossum" "Lasioglossum"
## [77] "Lasioglossum" "Macropis"    "Megachile"    "Megalopta"
## [81] "Melipona"     "Nomada"      "Nomia"        "Osmia"
## [85] "Osmia"        "Peponapis"   "Sphecodes"    "Stelis"
## [89] "Tetragonula"  "Tetragonula" "Tetragonula"  "Tetragonula"
## [93] "Tetragonula"
```

```
# Create grouped boxplot using ggplot2
```

```
len <- ggplot(tb, aes(x = Genus, y = Genome.Length..Mb., fill = Genus)) +
  geom_boxplot() +
  labs(x = "Genus", y = "Genome size (Mb)", title = "Genome size by Genus") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  guides(fill = FALSE)
```

```
## Warning: The '<scale>' argument of 'guides()' cannot be 'FALSE'. Use "none" instead as
## of ggplot2 3.3.4.
```

```
## This warning is displayed once every 8 hours.
```

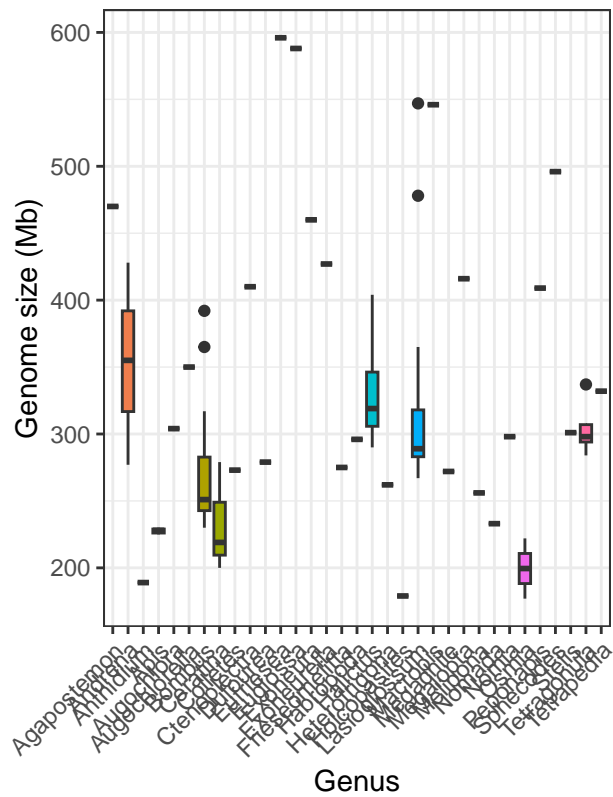
```
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

```
prot <- ggplot(na.omit(tb), aes(x = Genus, y = Annotated_Proteins, fill = Genus)) +
  geom_boxplot() +
  labs(x = "Genus", y = "Predicted Proteins", title = "Predicted Proteins by Genus") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  guides(fill = FALSE)
```

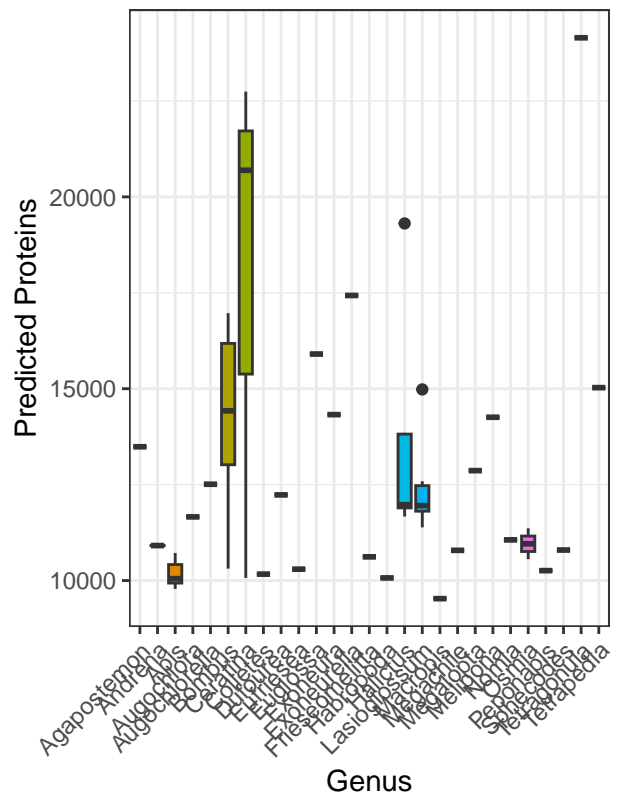
```
figure <- ggarrange(len, prot,
  labels = c("A", "B"),
  ncol = 2, nrow = 1)
```

```
figure
```

A Genome size by Genus



B Predicted Proteins by Genus



```
ggsave("genome_stats.png")
```

```
## Saving 6.5 x 4.5 in image
```