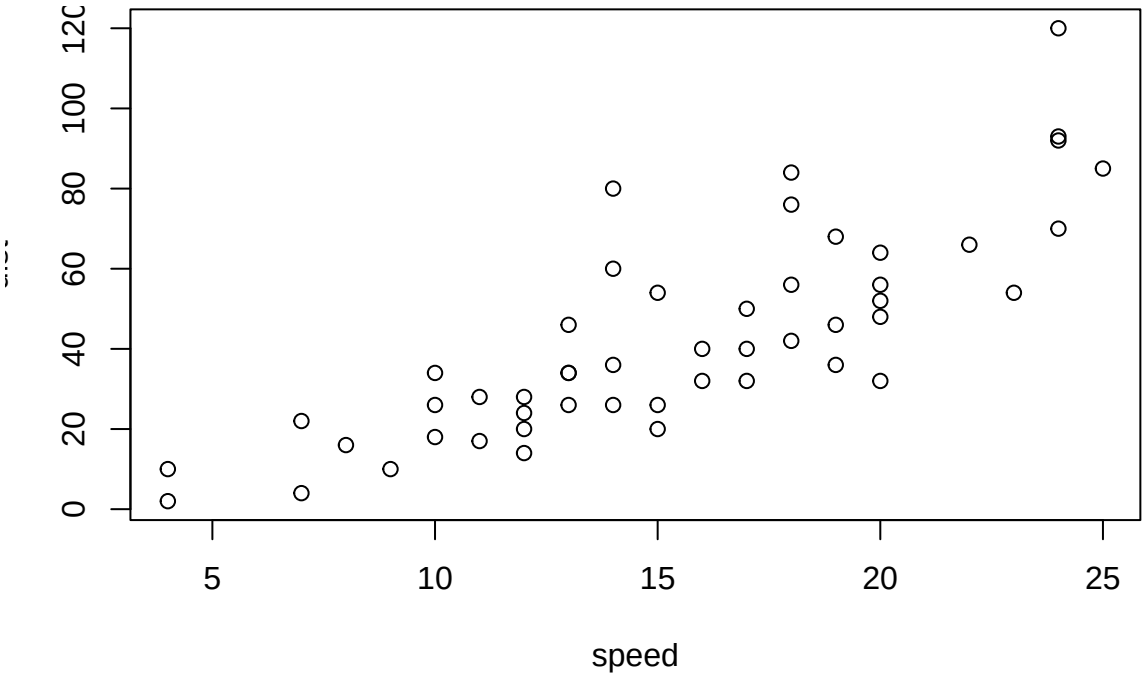
R Notebook

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

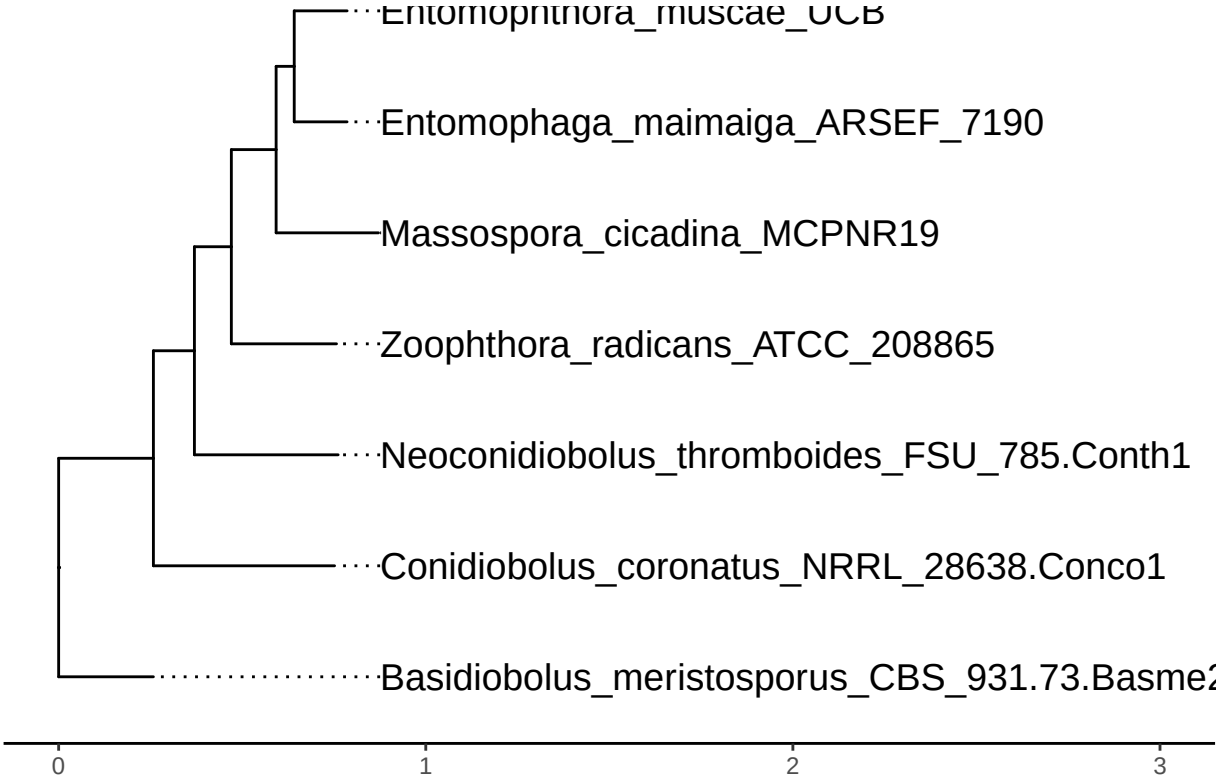
Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

```
plot(cars)
```



plot Figure 1

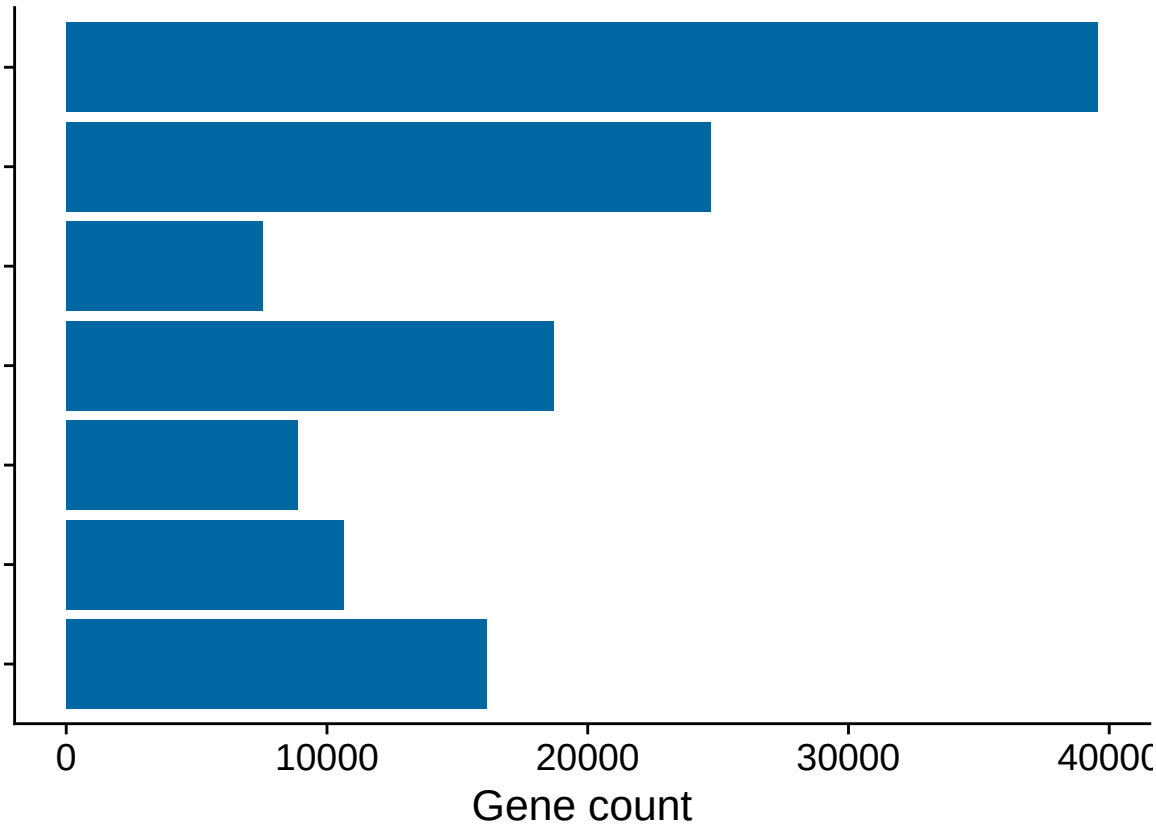
```
library(dplyr)
library(ggtree)
library(tidyverse)
library(ggsci)
library(scales)
library(NatParksPalettes)
library(wesanderson)
library(gridExtra)
library(cowplot)
tree <- read.tree("~/bigdata/Phylogeny/orthofinder/ento2/OrthoFinder/Results_Apr01/Species_Tree/Species

p<- ggtree(tree) + theme_tree2()+ xlim(0.0,3) +geom_tiplab(align = T, linesize = .5,size = 5)
print(p)
```



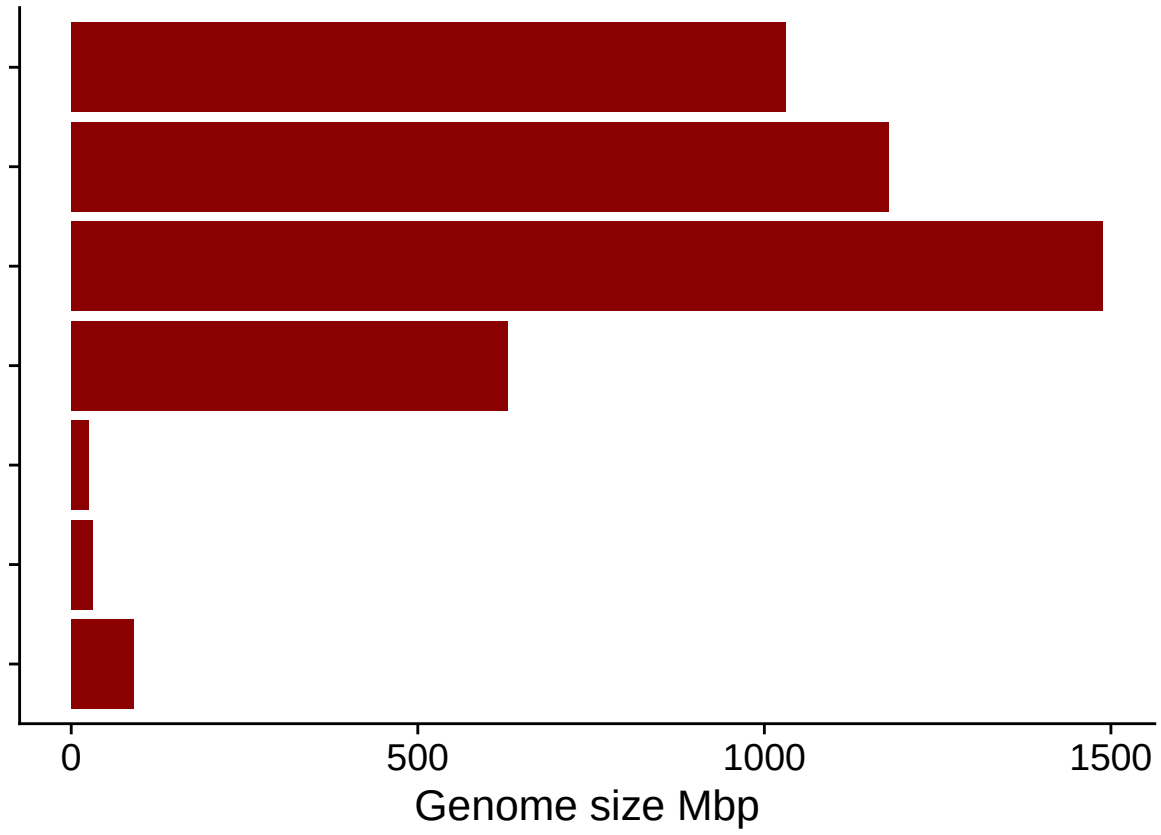

```
DF3 <- read_tsv("~/bigdata/Phylogeny/orthofinder/Taxa for phylogenetic tree building - Sheet6.tsv")
```

```
## Rows: 7 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: "\t"
## chr (1): species
## dbl (1): genecount
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

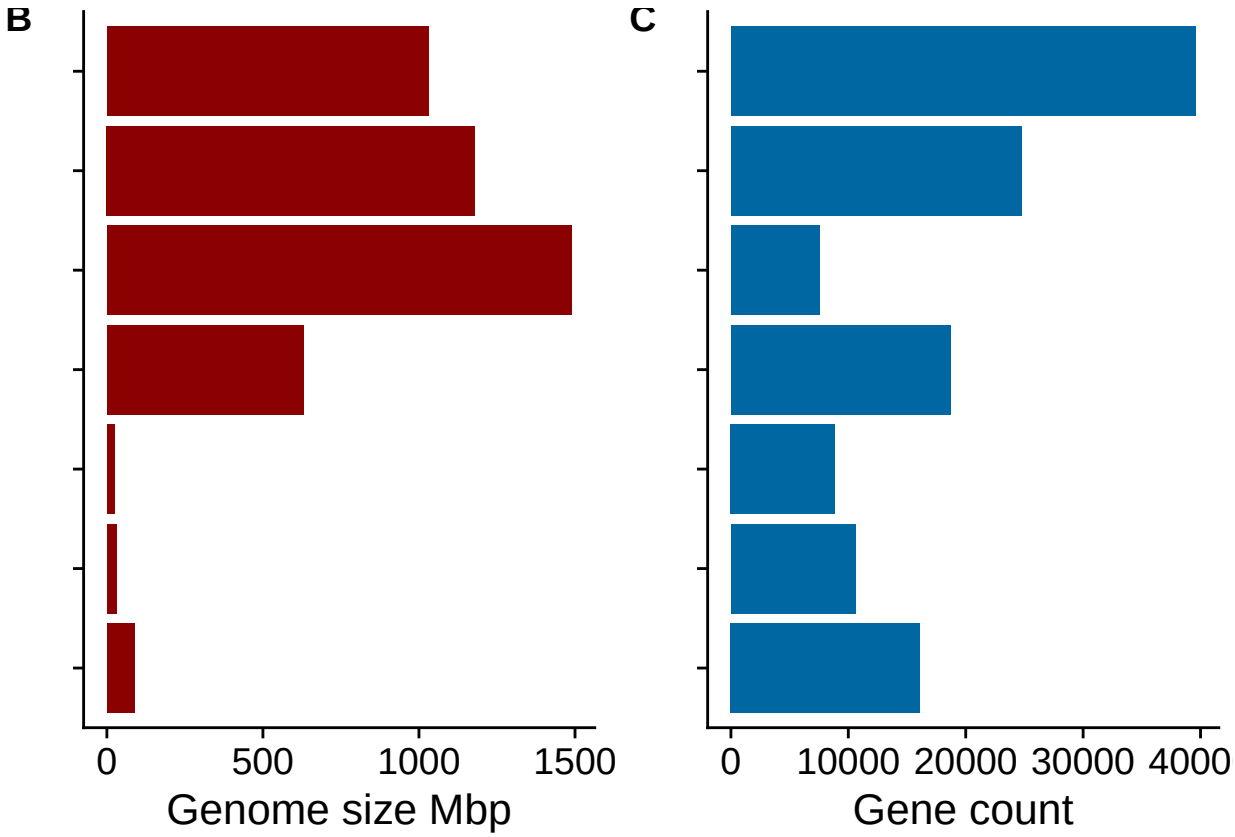
```
DF3$genomesize<-c(24.61,1031.19,1179.34,1488.88,629.39,31.71,89.49)
sp <- c("Entomophthora_muscae_UCB","Entomophaga_maimaiga_ARSEF_7190","Massospora_cicadina_MCPNR19","Zoop
sp <- rev(sp)

p3right<- DF3 %>% mutate( species=fct_relevel (species,sp))%>%
  ggplot(aes(x=species,y=genecount)) +
  geom_bar (stat = "identity", fill="#0067A2") +
  coord_flip()+
  theme_cowplot(16) +
  theme(axis.text.y  = element_text(size=0),
        legend.title = element_blank()) +
  xlab("") +
  ylab('Gene count')
 p3right
```




```
p3left<-DF3 %>% mutate(species=fct_relevel(species,sp)) %>%
  ggplot(aes(x=species,y=genomesize)) +
  geom_bar(stat="identity", fill = "darkred") +
  coord_flip()+
  theme_cowplot(16) +
 theme(axis.text.y  = element_text(size=0),
       legend.title = element_blank()) +
 xlab("") +
 ylab('Genome size Mbp')


p3left
```



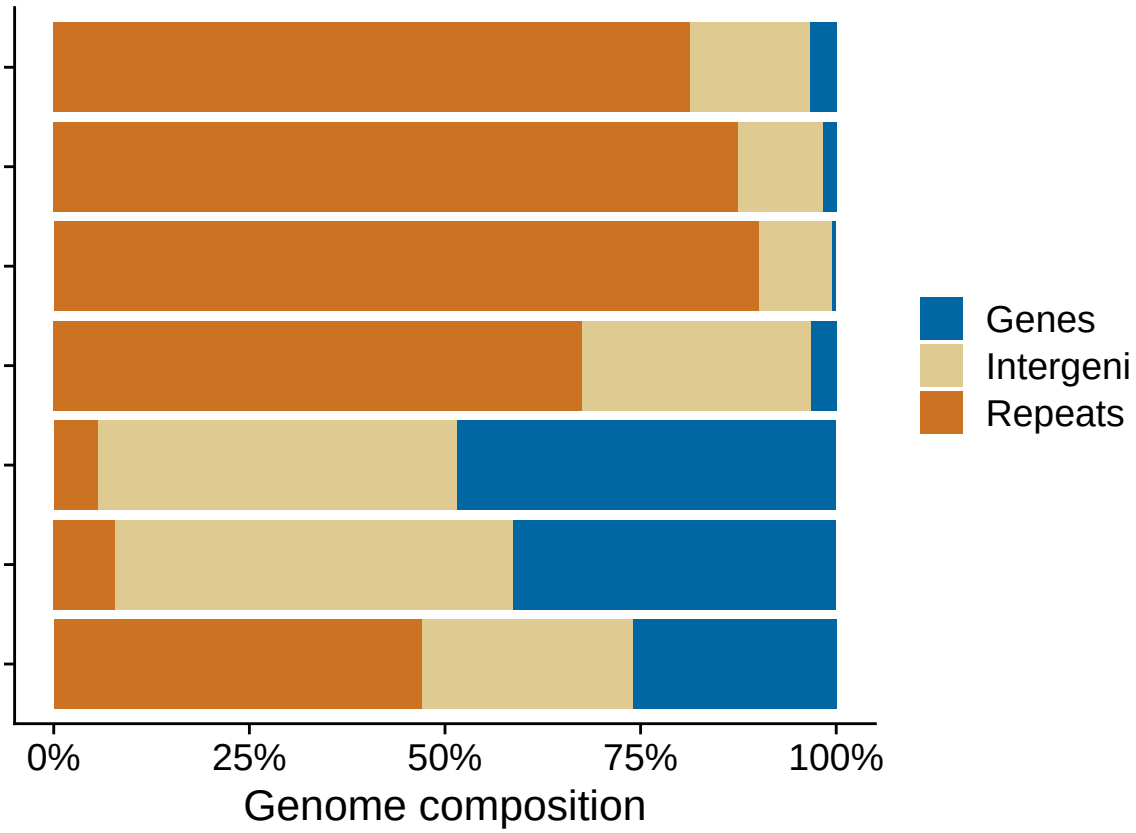

```
p3<-plot_grid(
 p3left,p3right,ncol = 2 , align = "h", axis="tb",labels = c("B","C")
)
p3
```



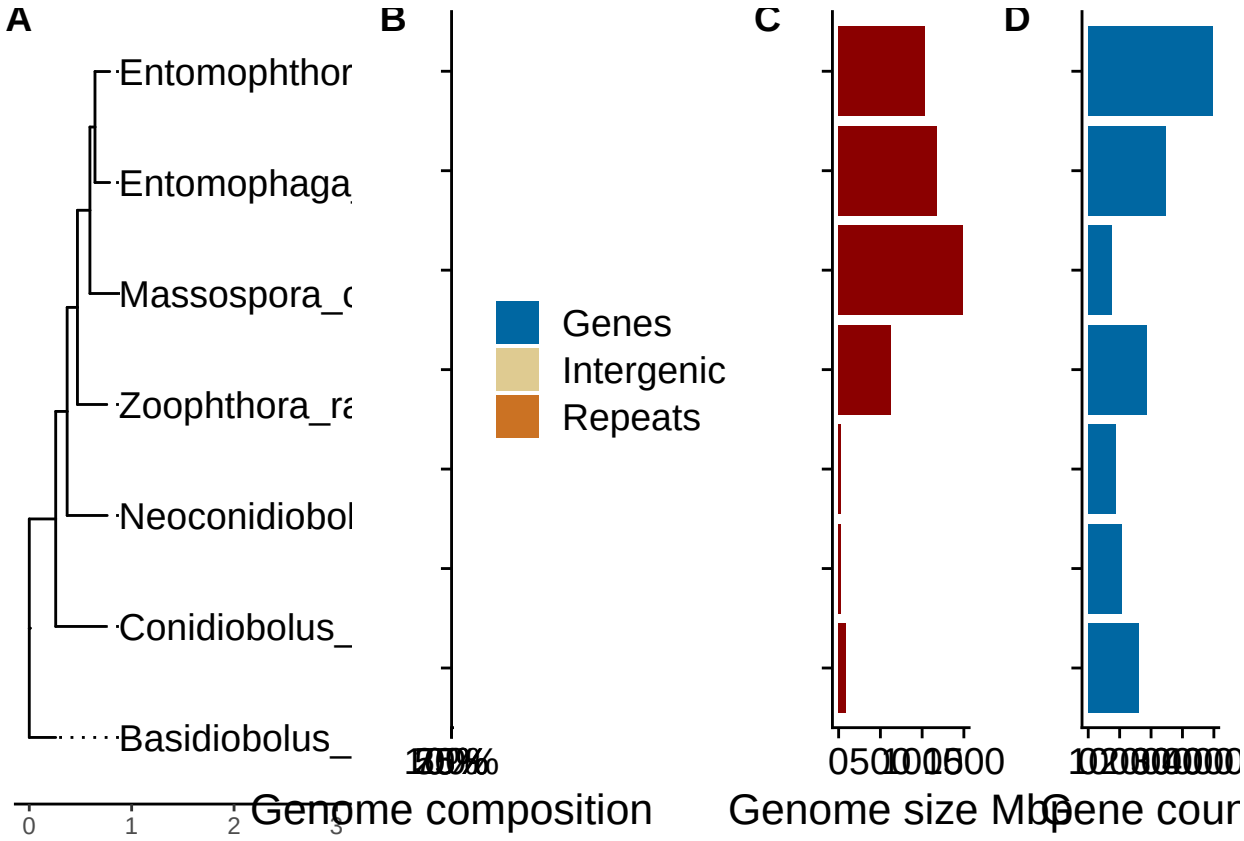

```
DF<-read_tsv("~/bigdata/Phylogeny/orthofinder/Taxa for phylogenetic tree building - Sheet4 (1).tsv",col
```

```
## Rows: 28 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: "\t"
## chr (2): species, type
## dbl (1): size
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
DF1 <- DF %>% filter(type != "Genome")

sp <- c("Entomophthora_muscae_UCB","Entomophaga_maimaiga_ARSEF_7190","Massospora_cicadina_MCPNR19","Zoo
sp <- rev(sp)


p2<-DF1 %>% mutate ( species=fct_relevel (species,sp)) %>%
  ggplot(aes(x=species,y=size*100,fill=factor(type,levels=c("Genes","Intergenic","Repeats")))) + geom_ba
  scale_fill_manual(values = natparks.pals("Yellowstone",3,override.order = F)) +
  theme_cowplot(16)+
  theme(axis.text.y  = element_text(size=0),
        legend.title = element_blank()) +
  xlab("") +
  ylab("Genome composition") +
  scale_y_continuous(labels = scales::percent_format(scale = 100))
print(p2)
```



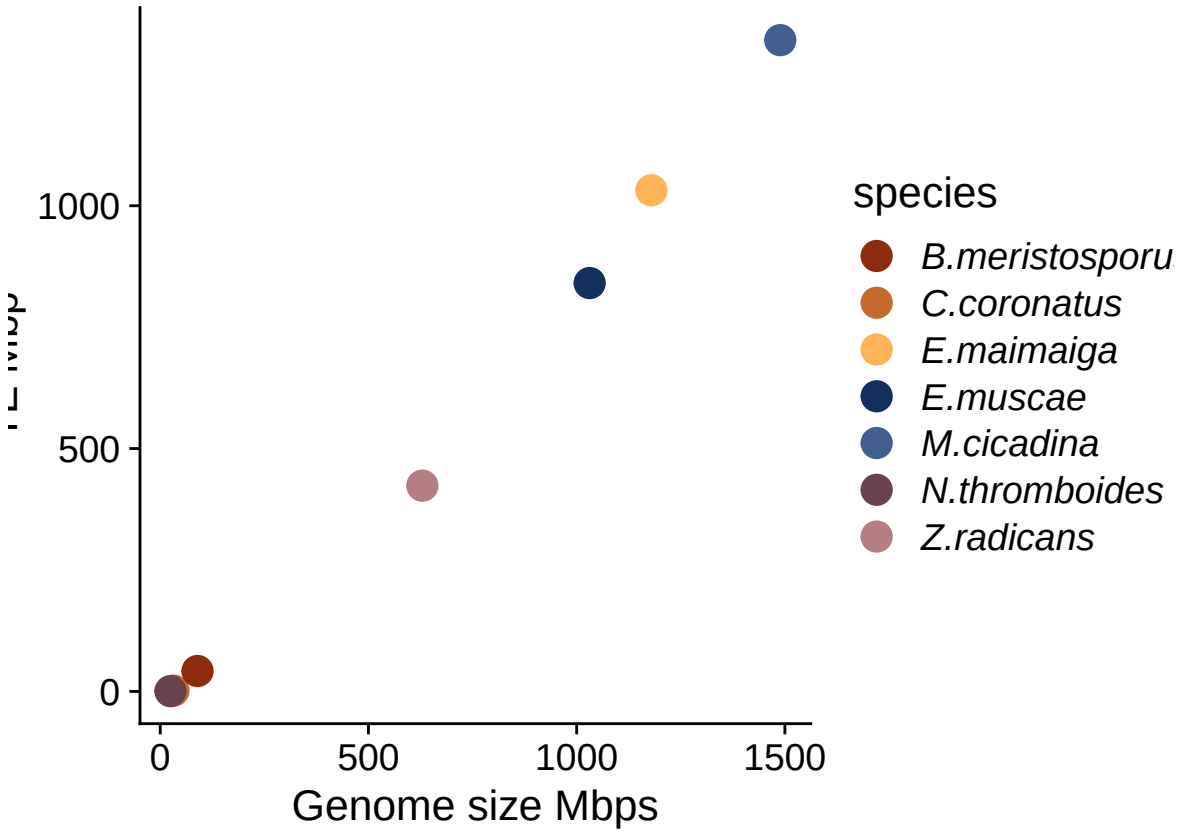

```
upperfig1<-plot_grid(p,p2,p3left,p3right,nrow = 1,rel_widths = c(1.5,1.5,1,1),labels = c("A","B","C","D
upperfig1
```




```
DF2<-read_tsv("~/bigdata/Phylogeny/orthofinder/TEVSgenome.tsv",col_names = T)
```

```
## Rows: 7 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: "\t"
## chr (1): species
## dbl (2): Genome, TE
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
DF2<-DF2%>%mutate(Genome=Genome/1000000,TE=TE/1000000)
fig1e<-DF2 %>% ggplot(aes(x=Genome,y=TE,color=species)) +
  geom_point(size =5) +
  theme_cowplot(16)+
  scale_color_manual(values = natparks.pals("DeathValley"))+
  xlab("Genome size Mbps") +
  ylab("TE Mbp")+
  theme(legend.text = element_text(face = "italic"))



print(fig1e)
```



head to file genomecompositionFeb24 to import composition plots

```
library(tidyverse)
library(readr)
library(dplyr)
library(ggsci)
library(gridExtra)
library(ggplot2)
library(RColorBrewer)
library(NatParksPalettes)
total <- read_tsv("~/bigdata/TE_composition-EDTA/tables/TE_compositionFeb2024_bpmasked.tsv",col_names =
```

```
## Rows: 20 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: "\t"
## chr (2): Class, species
## dbl (1): bpmasked
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

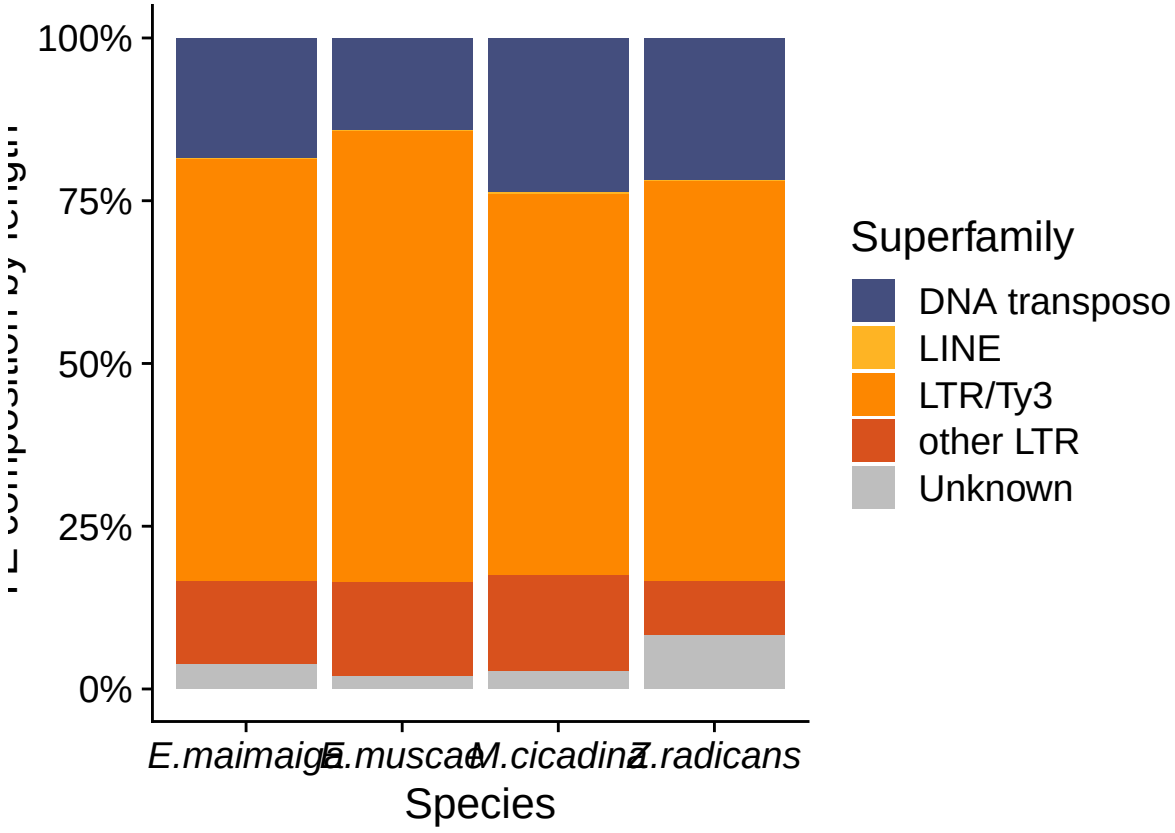
```
colors<-natparks.pals("Acadia")
for (i in 1:length(colors)) {
  cat(sprintf("Color %d: %s\n", i, colors[i]))
}
```

```
## Color 1: #212E52
## Color 2: #444E7E
```

```
## Color 3: #8087AA
## Color 4: #B7ABBC
## Color 5: #F9ECE8
## Color 6: #FCC893
## Color 7: #FEB424
## Color 8: #FD8700
## Color 9: #D8511D
```

```
my_palette <- c("#444E7E", "#FEB424","#FD8700", "#D8511D", "grey")
```

```
fig1f<-total %>% ggplot(aes(x=species, y=bpmasked, fill=Class)) +
  geom_bar(position = "fill", stat = "identity") +
  labs(fill = "Superfamily") +
  theme_cowplot(16)+
  scale_fill_manual(values = my_palette)+
    theme(axis.text.x = element_text(face = 3))+
  ylab("TE composition by length") +
  scale_y_continuous(labels = scales::percent_format(scale = 100))+
  xlab("Species")
print(fig1f)
```



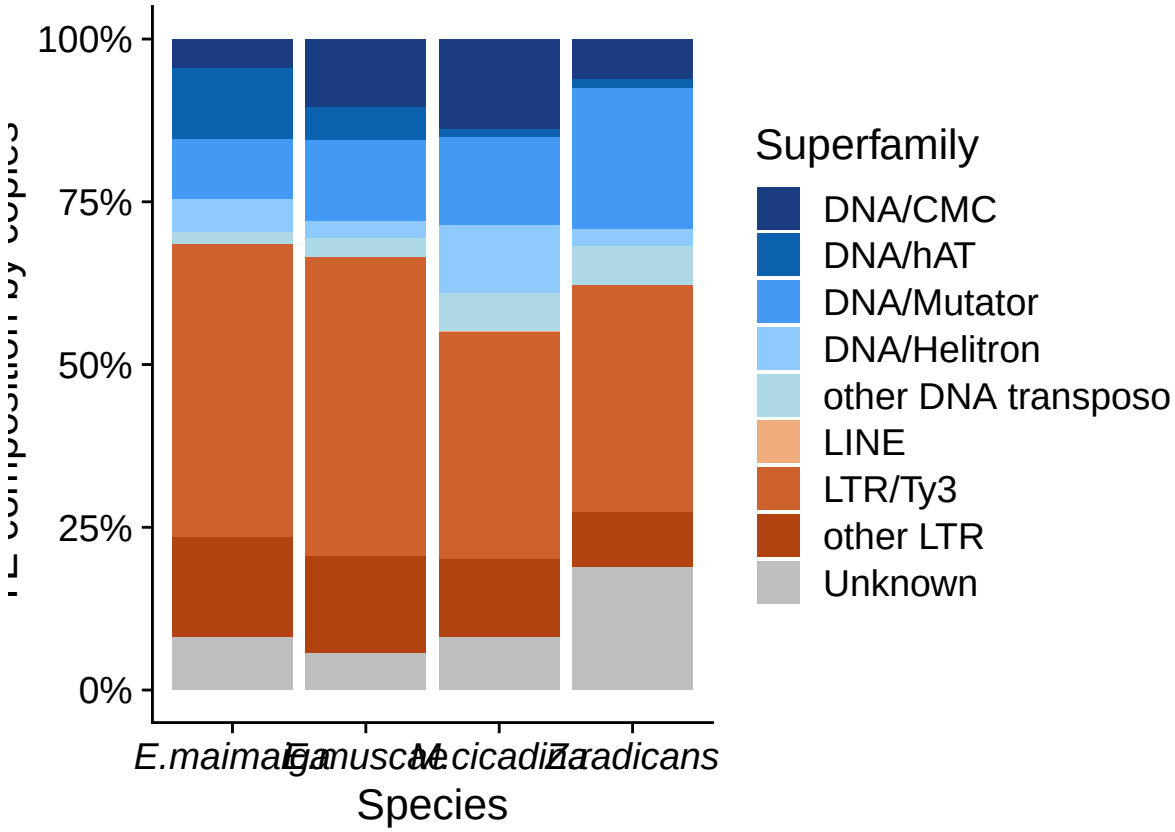

```
totalnumber <- read_tsv("~/bigdata/TE_composition-EDTA/tables/TE_compositionFeb2024_count.tsv",col_names
```

```
## Rows: 36 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: "\t"
## chr (2): Class, species
## dbl (1): count
```

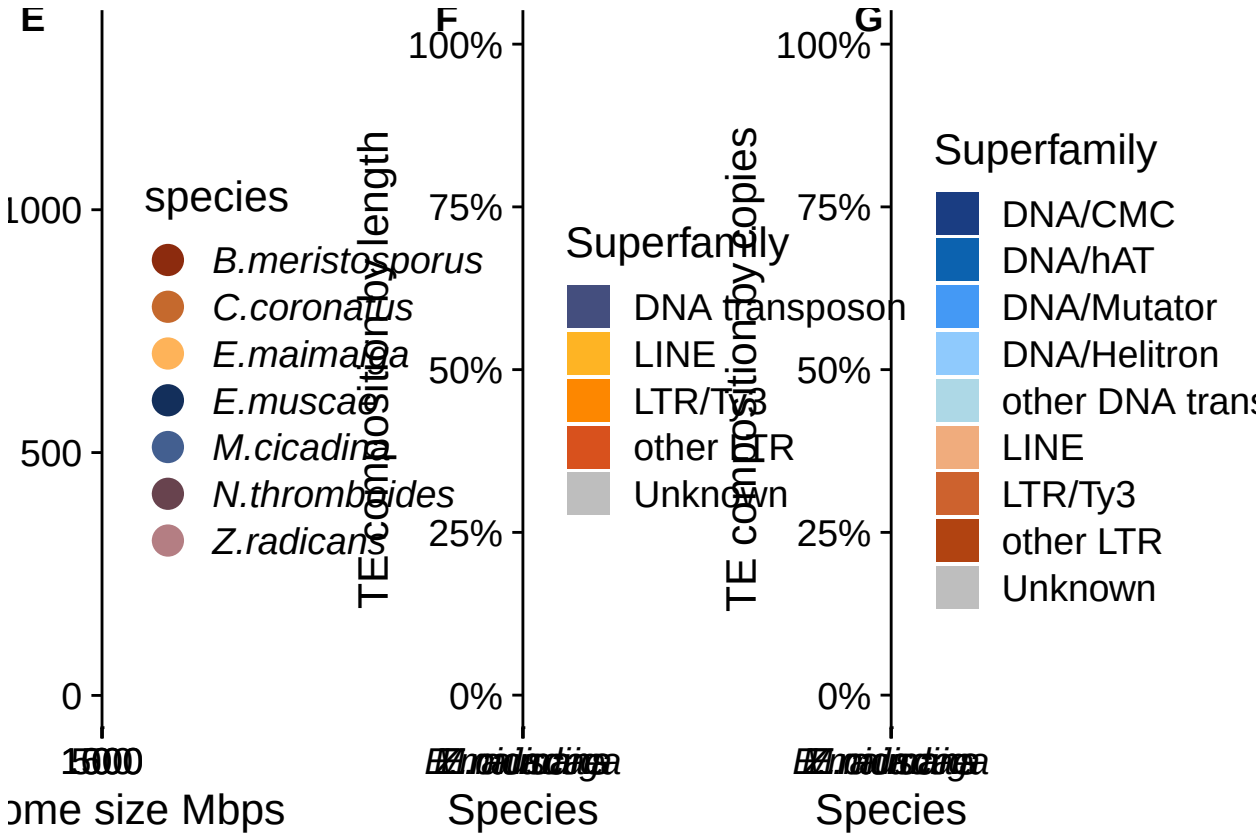
```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
table1 <- totalnumber %>% group_by(species) %>%
  summarize(total=sum(count))

cl <- c("Unknown", "other LTR", "LTR/Ty3","LINE","other DNA transposon","DNA/Helitron","DNA/Mutator","D
cl <- rev(cl)
```

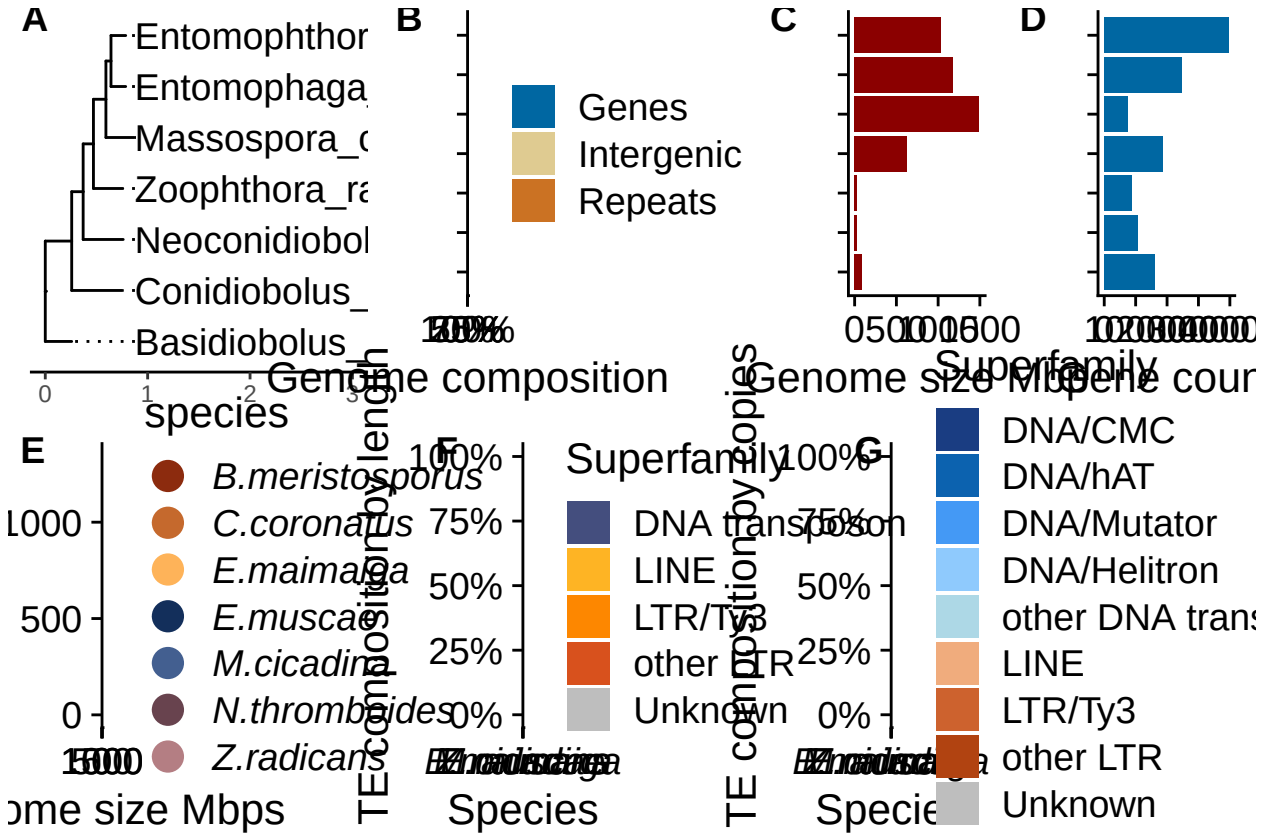
```
my_palette2<-c("#1A3D82","#0C62AF","#4499F5","#8FCAFD","lightblue","#F0AC7D","#CD622E","#B14311","grey"

fig1g<-totalnumber %>%mutate ( Class=fct_relevel (Class,cl)) %>% ggplot(aes(x=species, y=count, fill=Cla
  geom_bar(position = "fill", stat = "identity") +
  theme_cowplot(16)+
  labs(fill = "Superfamily")+
  scale_fill_manual(values=my_palette2)+
  scale_y_continuous(labels = scales::percent_format(scale = 100))+
  theme(axis.text.x = element_text(face = 3))+
  ylab("TE composition by copies") +
  xlab("Species")
print(fig1g)
```




```
bottomfig1<-plot_grid(fig1e,fig1f,fig1g, labels = c("E","F","G"),nrow = 1)
print(bottomfig1)
```




```
fig1 <- plot_grid(
  upperfig1,bottomfig1,
  nrow = 2)
print(fig1)
```




```
ggsave(filename = "~/bigdata/Phylogeny/orthofinder/fig1.pdf",
       plot = fig1,
       units = "in",
       width = 23,
       height = 15,
       dpi = 500)
```