ggtree_example

Susan Gogolski

2025-04-02

```r
library(ggplot2)
library(viridis)
```

```
## Loading required package: viridisLite
```

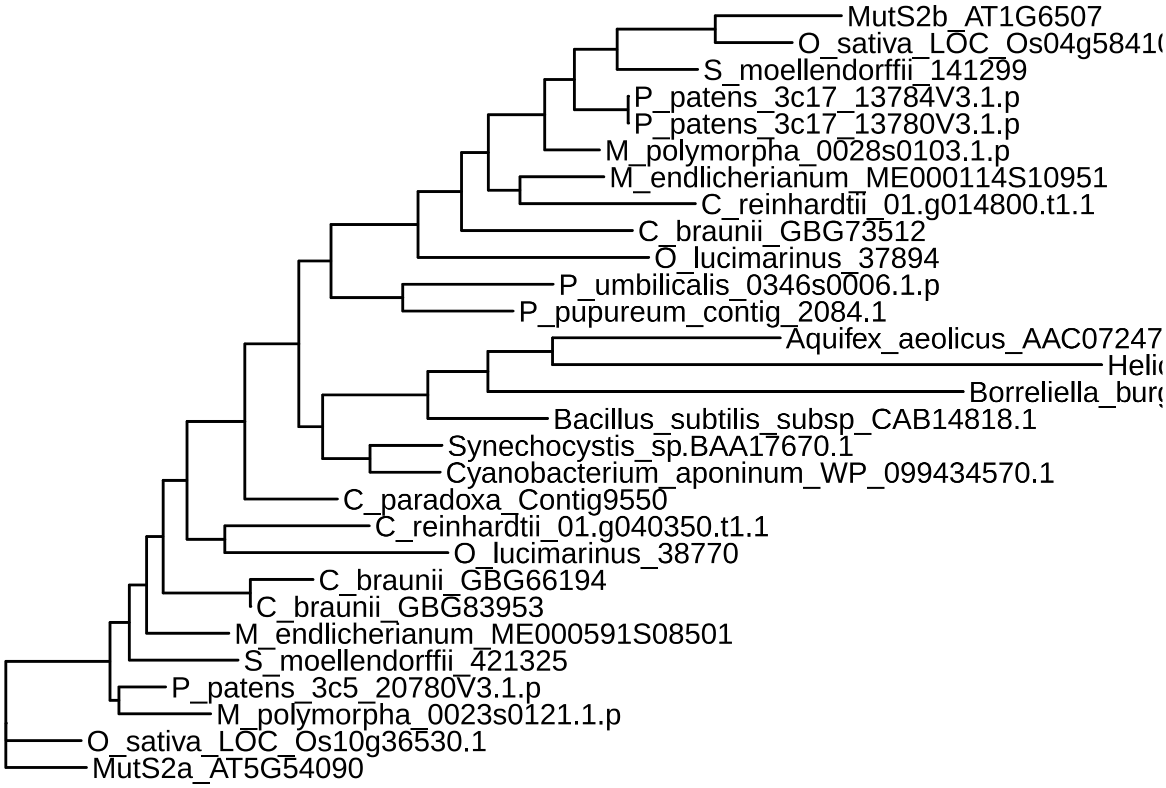
```r
if (!requireNamespace("BiocManager", quietly = TRUE)){
  install.packages("BiocManager")
}

if (!requireNamespace("ggtree", quietly = TRUE)){
  BiocManager::install("ggtree")
}
library(ggtree)
```

```
## ggtree v3.14.0 Learn more at https://yulab-smu.top/contribution-tree-data/
##
## Please cite:
##
## Guangchuang Yu, Tommy Tsan-Yuk Lam, Huachen Zhu, Yi Guan. Two methods
## for mapping and visualizing associated data on phylogeny using ggtree.
## Molecular Biology and Evolution. 2018, 35(12):3041-3043.
## doi:10.1093/molbev/msy194
```

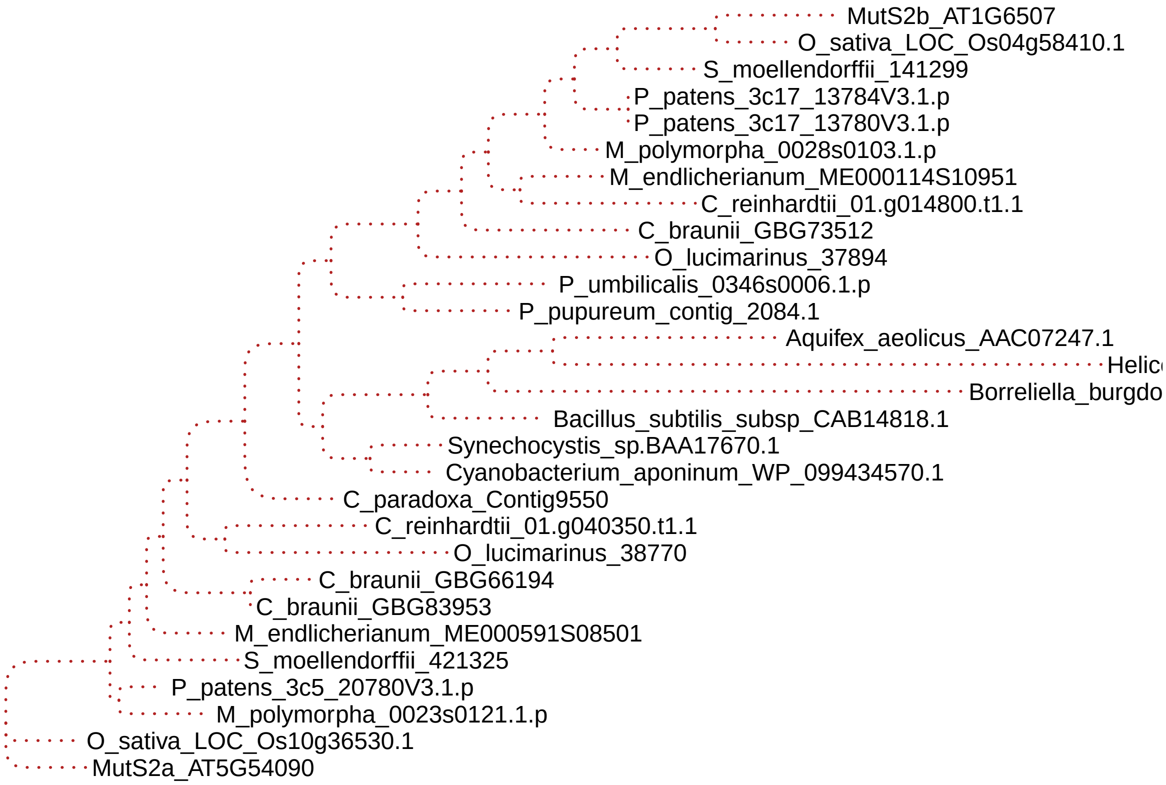
```r
tree = ggtree::read.tree(file = "tree.nwk")
tree
```

```
##
## Phylogenetic tree with 29 tips and 27 internal nodes.
##
## Tip labels:
##   MutS2a_AT5G54090, S_moellendorffii_421325, C_paradoxa_Contig9550, Cyanobacterium_aponinum_WP_099434
##
## Unrooted; includes branch length(s).
```

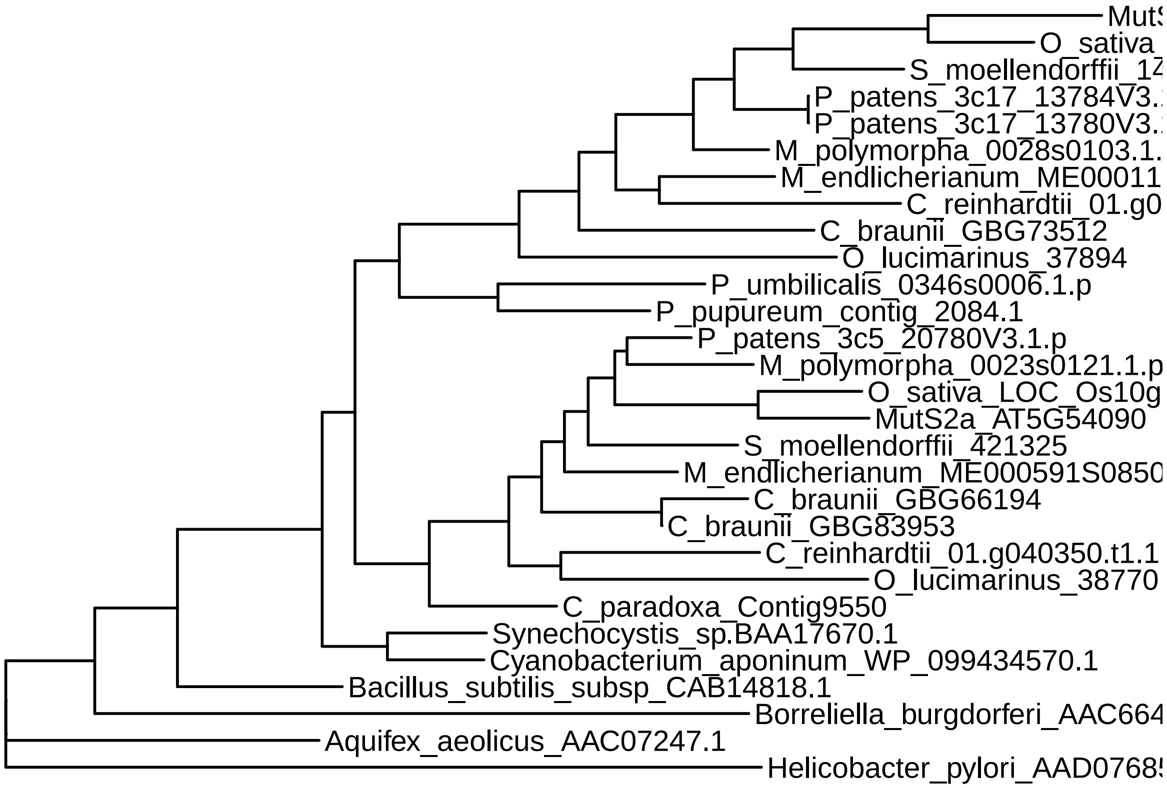
```r
ggplot(tree) + geom_tree() +
  geom_tiplab() +
  theme_tree()
```



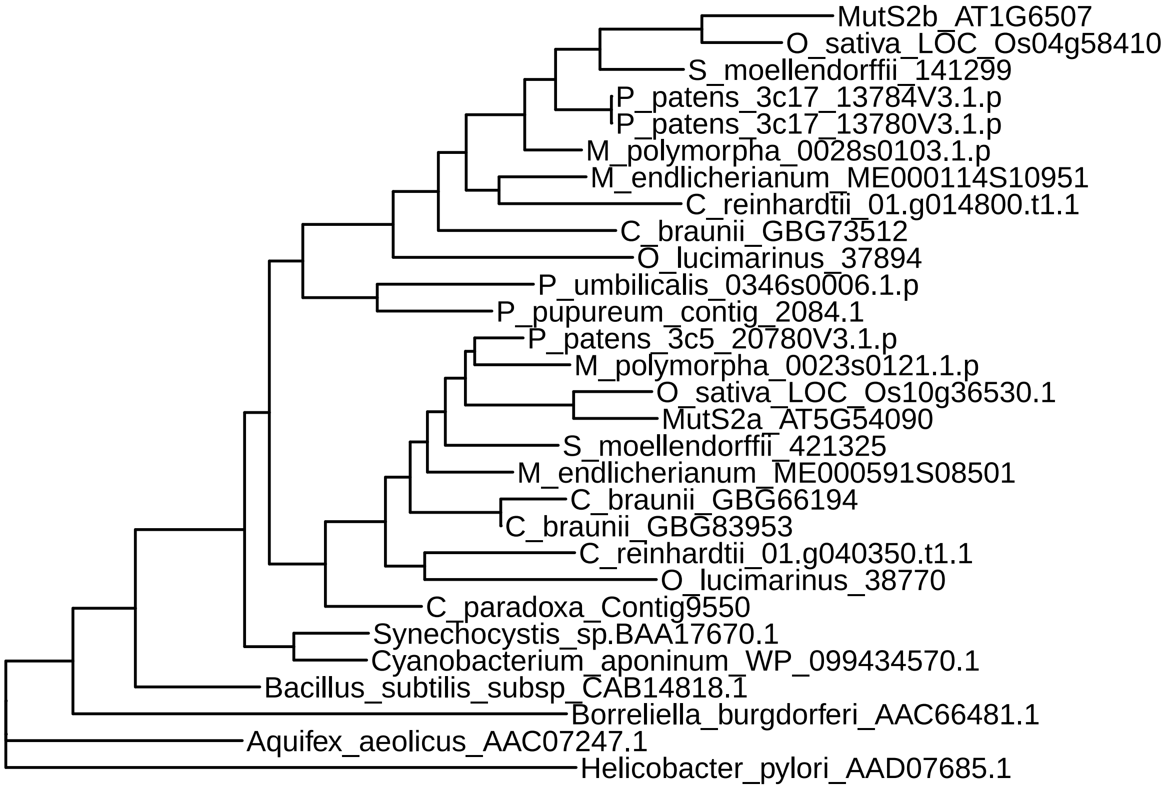
```r
ggplot(tree) +
  geom_tree(color="firebrick", size=0.5, linetype="dotted", layout="roundrect") +
  geom_tiplab(size=3) +
  theme_tree()
```



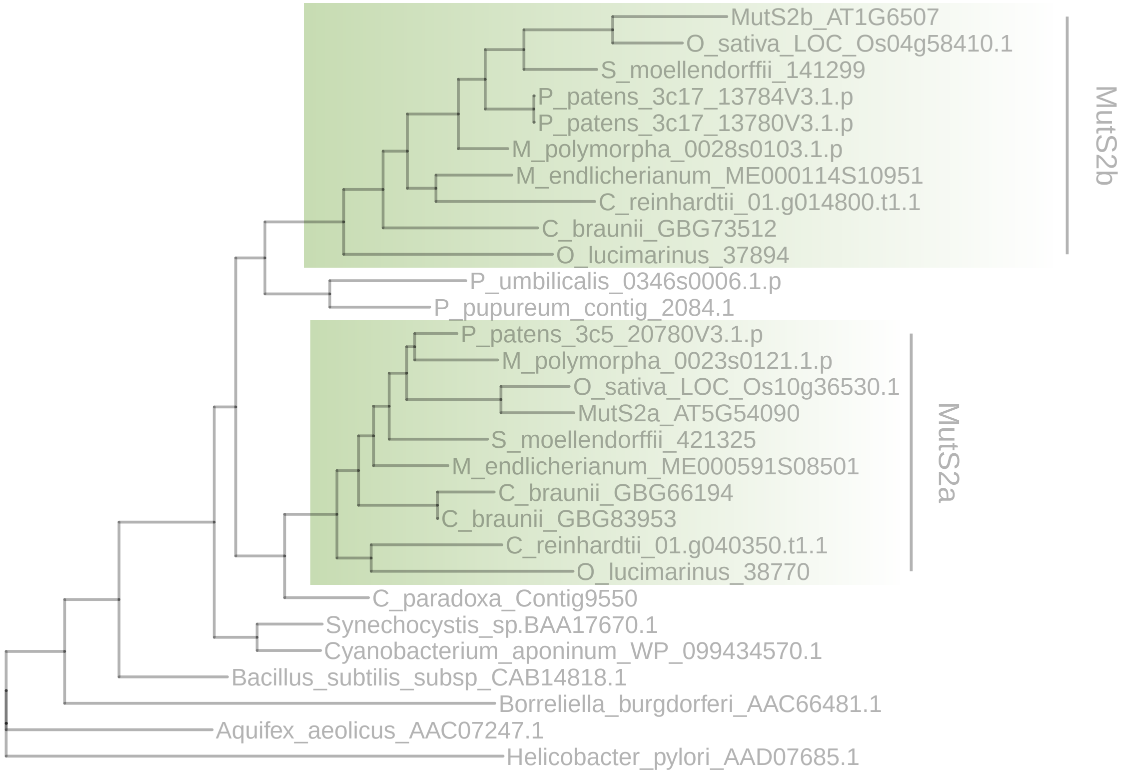
```r
tree_rooted = ape::root(tree, outgroup = "Helicobacter_pylori_AAD07685.1", resolve.root=TRUE)

ggplot(tree_rooted) +
  geom_tree() +
  geom_tiplab() +
  theme_tree()
```



```r
ggplot(tree_rooted) +
  geom_tree() +
  geom_tiplab() +
  theme_tree() +
  hexpand(0.25)
```



```
ggplot(tree_rooted) +
  geom_tree() +
  geom_tiplab(size=3) +
  theme_tree() +
  hexpand(0.25) +
  geom_hilight(node=36, fill="chartreuse4", alpha=.3, type = "gradient", to.bottom=TRUE, extend=1) +
  geom_hilight(node=44, fill="chartreuse4", alpha=.3, type = "gradient", to.bottom=TRUE, extend=1) +
  geom_cladelab(node=36, label = "MutS2a", angle=270, hjust = "center", offset=1, offset.text=0.1) +
  geom_cladelab(node=44, label = "MutS2b", angle=270, hjust = "center", offset=1, offset.text=0.1)
```

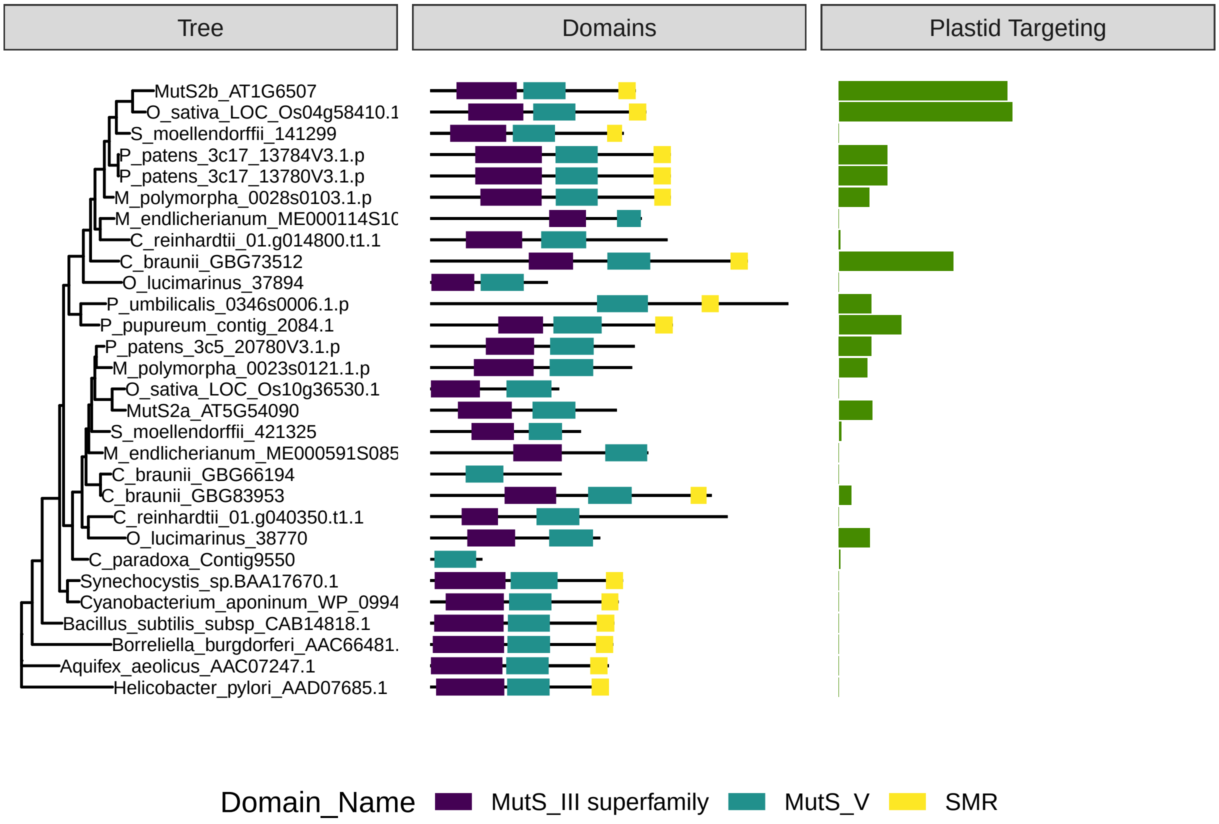


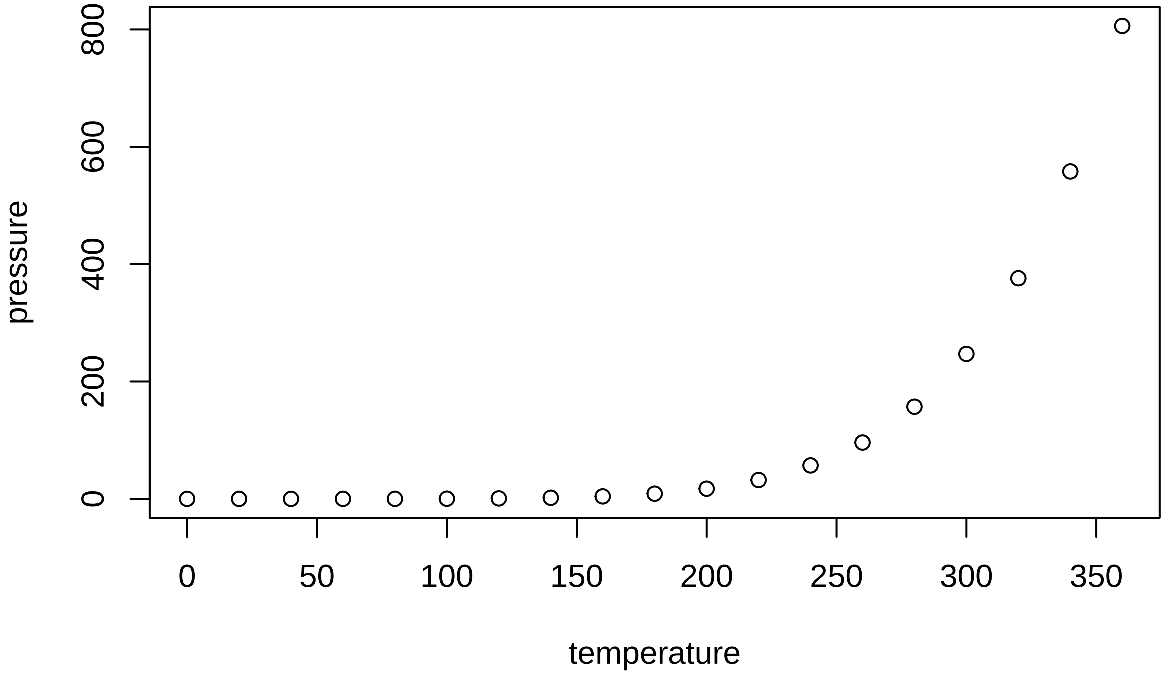
```r
gene_data = read.delim("gene_data.txt")
gene_data[1:5,]
```

```
##                Newick_label Domain_Start Domain_End          Domain_Name
## 1          MutS2a_AT5G54090          118        348 MutS_III superfamily
## 2          MutS2a_AT5G54090          436        619               MutS_V
## 3 O_sativa_LOC_Os10g36530.1            3        212 MutS_III superfamily
## 4 O_sativa_LOC_Os10g36530.1          325        517               MutS_V
## 5        C_braunii_GBG66194          151        312               MutS_V
##   Seq_Length Plastid_Targeting
## 1        796              0.28
## 2        796              0.28
## 3        550              0.00
## 4        550              0.00
## 5        561              0.00
```

```r
#store the tree
p1 = ggplot(tree_rooted, aes(x,y)) +
  geom_tree() +
  geom_tiplab(size=2.5) +
  hexpand(1.5) +
  theme_tree()

#add a facet with lines corresponding to the length of each protein
p2 = facet_plot(p1, panel = "Domains", data = gene_data, geom = geom_segment, mapping = aes(x=0, xend=S

#add the domains on top of those lines
p3 = facet_plot(p2, panel = "Domains", data = gene_data, geom=geom_segment, aes(x=Domain_Start, xend=Do
```

```r
  theme(legend.position = "bottom") +
  scale_color_viridis(discrete = TRUE)

#add another facet with a bar graph showing the probability of plastid targeting
p4 = facet_plot(p3, panel = "Plastid Targeting", data = gene_data, geom=geom_col, aes(x=Plastid_Targeti
  theme(legend.position = "bottom")

p4
```





Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.