

# Ggtree: 用于系统发育树及 相关数据存储与可视化的 数据结构

徐双斌, 李林, 罗晓, 陈玫君, 唐文丽, 詹历, 戴泽翰  
林讚育, 管轶, 余光创

南方医科大学基础医学院生物信息学系  
香港大学公共卫生学院新发传染病国家重点实验室  
汕头大学病毒学联合研究所



Shuangbin Xu, Lin Li, Xiao Luo, Meijun Chen, Wenli Tang, Li Zhan, Zehan Dai, Tommy T. Lam, Yi Guan, Guangchuang Yu. 2022. “Ggtree: A serialized data object for visualization of a phylogenetic tree and annotation data.” *iMeta*. e56. <https://doi.org/10.1002/imt2.56>

# 正文

## 系统发育树+相关数据信息



Ecology



Epidemiology



Microbiome



对于系统发育数据的重新利用有助于在许多学科中对系统发育信息进行综合比较分析

## 相关问题

- 大约60%的已发表系统发育数据在科学社区中已永远丢失[1]。造成这一情况的原因之一是系统发育数据的发表共享通常是以静态图像为主，缺乏可互通的共享文件[2]。
- 尽管用于系统发育树的可视化以及注释的工具正在逐渐增加，但是这些工具的目的仍然是为了生成可发表的系统发育树图，其中还涉及选择注释数据（如bootstrap值）并将其注释到系统发育树上（比如作为文本标签或者分支颜色）。而这一过程也往往是单向无法逆转地产生一个静态图形，使得一些基础信息无法重用。

## 潜在的解决方法

需要从这种生成静态图形的单向目的转换为将系统发育树，相关数据以及可视化指令组成统一的数据对象同时又可渲染成可视化图形的模式。

[1] <https://doi.org/10.1371/journal.pone.0110268>

[2] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4073804/>

# 正文

ggtree数据结构



系统发育树 + 相关数据信息

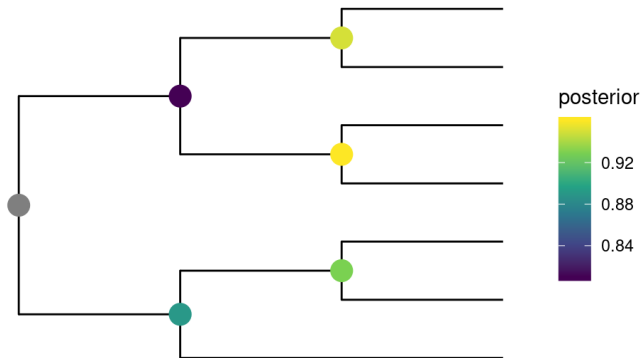
如何构建这个数据结构



```
# Loading the required packages
pacman::p_load(tidytree, treeio, ggplot2, ggtree)
# The url of phylogenetic tree and associated data,
# which can be replace user own files.
url <- paste0("https://raw.githubusercontent.com/TreeViz/",
"metastyle/master/design/viz_targets_exercise/")
# parsing the phylogenetic tree files with
# the functions of treeio package, it will generate
# phylo or treedata object.
x <- read.tree(paste0(url, "tree_boots.nwk"))
# reading the associated data
d <- read.csv(paste0(url, "inode_data.csv"))
# constructing the ggtree object (using ggtree function) and
# adding associated data to the object (using %<+% function)
p <- ggtree(x) %<+% d +
# annotating tree with the posterior (in this example) or other data
  geom_nodepoint(aes(colour = posterior), size = 5) +
# adjust the color of the data point annotated.
  scale_color_viridis_c() +
# adjusting the theme of the object.
  theme(legend.position = 'right')
```

该对象可以被渲染为静态图片

→ `print(p)`



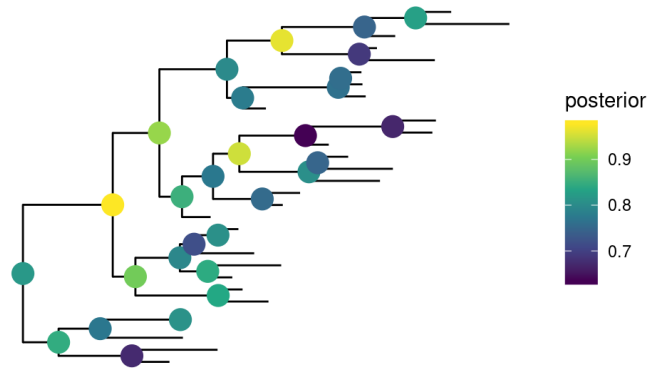
# 正文

从ggtree对象中提取出系统发育信息

```
## extract tree from graphic object
tree <- as.treedata(p)
## associated data is included in the tree object
get.fields(tree)
## [1] "vernacularName" "infoURL" "rank" "bootstrap"
## [5] "posterior"
## convert graphic object to Newick text
## tree can be exported with associated data into
## a single file using write.beast
write.tree(as.phylo(p))
## [1]
"(((Rangifer_tarandus:1,Cervus_elaphus:1)Cervidae:1,(Bos_taurus:1,Ovis_orient
alis:1)Bovidae:1)Artiodactyla:1,(Suricata_suricatta:2,(Cystophora_cristata:1,
Mephitis_mephitis:1)Caniformia:1)Carnivora:1)Mammalia;"
```

ggtree对象也可以被用来给新的树(phylo或者treedata)对象进行可视化，这与Microsoft的格式刷类似。

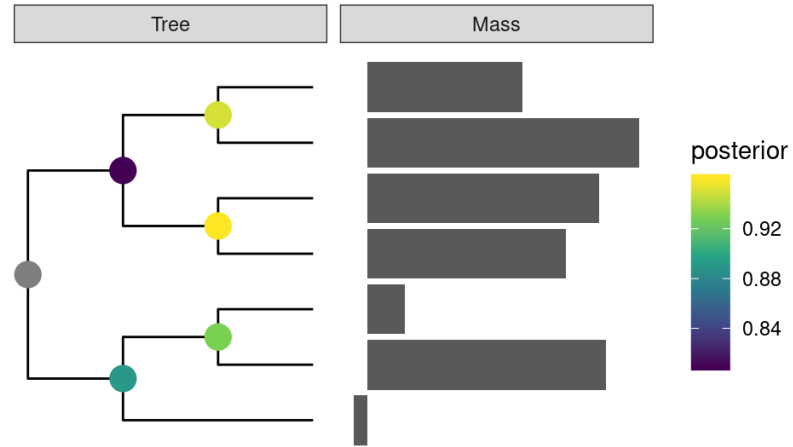
```
y <- treedata(phylo = rtree(30),
              data = tibble(node = 31:59,
                             posterior = rnorm(29, 0.8, .1)))
p %<% y
```



# 正文

相关的系统发育数据可用过  
facet\_plot 与ggtree对象进行连接

```
info <- read.csv(paste0(url, "tip_data.csv"))
p2 <- facet_plot(p, data = info[, c(1,7,8)],
  geom = geom_col,
  mapping = aes(x=log(mass_in_kg)),
  orientation = 'y', panel = 'Mass')
```



而这些相关数据可以通过facet\_data  
来重新提取出来。

```
facet_data(p2, 'Mass')
```

##	label	mass_in_kg	trophic_habit
## 1	Bos_taurus	618.64	herbivore
## 2	Cervus_elaphus	240.87	herbivore
## 3	Cystophora_cristata	278.90	omnivore
## 4	Mephitis_mephitis	2.40	omnivore
## 5	Ovis_orientalis	39.10	herbivore
## 6	Rangifer_tarandus	109.09	herbivore
## 7	Suricata_suricatta	0.73	carnivore

# 总结

- 系统发育树与形式多样的相关数据可以存储在同一个ggtree图形对象中从而提高了系统发育数据的可重复性与可重用性。
- 系统发育树与相关数据可以从ggtree对象中提取出来，并重新分析以辅助各个学科进行系统发育信息的综合比较研究。
- Ggtree图形对象可以被渲染成静态图像，并且保存在该对象中的可视化指令可以类似于Microsoft Word 格式刷操作对不同的树对象进行可视化。

Shuangbin Xu, Lin Li, Xiao Luo, Meijun Chen, Wenli Tang, Li Zhan, Zehan Dai, Tommy T. Lam, Yi Guan, Guangchuang Yu. 2022. “Ggtree: A serialized data object for visualization of a phylogenetic tree and annotation data.” *iMeta*. e56. <https://doi.org/10.1002/imt2.56>



“iMeta”是由威立、肠菌分会和本领域数百位华人科学家合作出版的开放获取期刊，主编由中科院微生物所刘双江研究员和荷兰格罗宁根大学傅静远教授共同担任。目的是发表原创研究、方法和综述以促进宏基因组学、微生物组和生物信息学发展。目标是发表前10%(IF > 15)的高影响力论文。期刊特色包括视频投稿、可重复分析、图片打磨、青年编委、前3年免出版费、50万用户的社交媒体宣传等。2022年的三月、六月和九月期已正式在线出版发行!



主页: <http://www.imeta.science>

出版社: <https://wileyonlinelibrary.com/journal/imeta>



投稿: <https://mc.manuscriptcentral.com/imeta>



[office@imeta.science](mailto:office@imeta.science)



[iMeta](#)

[宣传片](#)

