

# Data integration and visualization of phylogenetic trees

Guangchuang YU

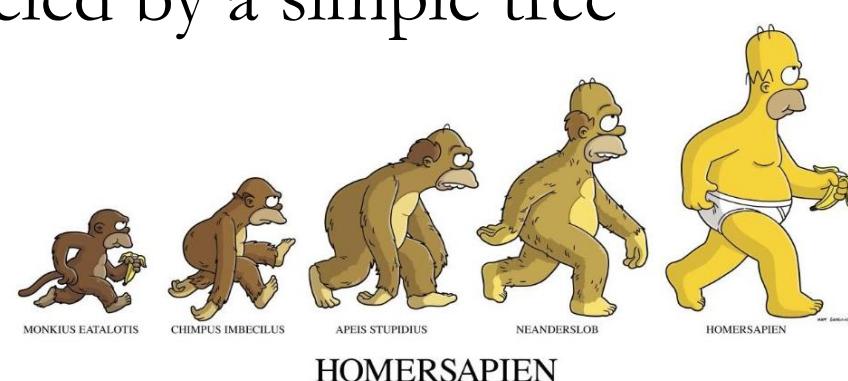
Southern Medical University

Guangzhou, China

# What is a phylogenetic tree?

A model of evolutionary history

- Understanding life **through time** (over long periods of past time)
- The connections between all groups of organisms are understood by **ancestor-descendant** relationships
- A phylogenetic tree is **one possible** representation of relationships
- Evolutionary events like hybridization, horizontal gene transfer, recombination and reassortment can't be modeled by a simple tree

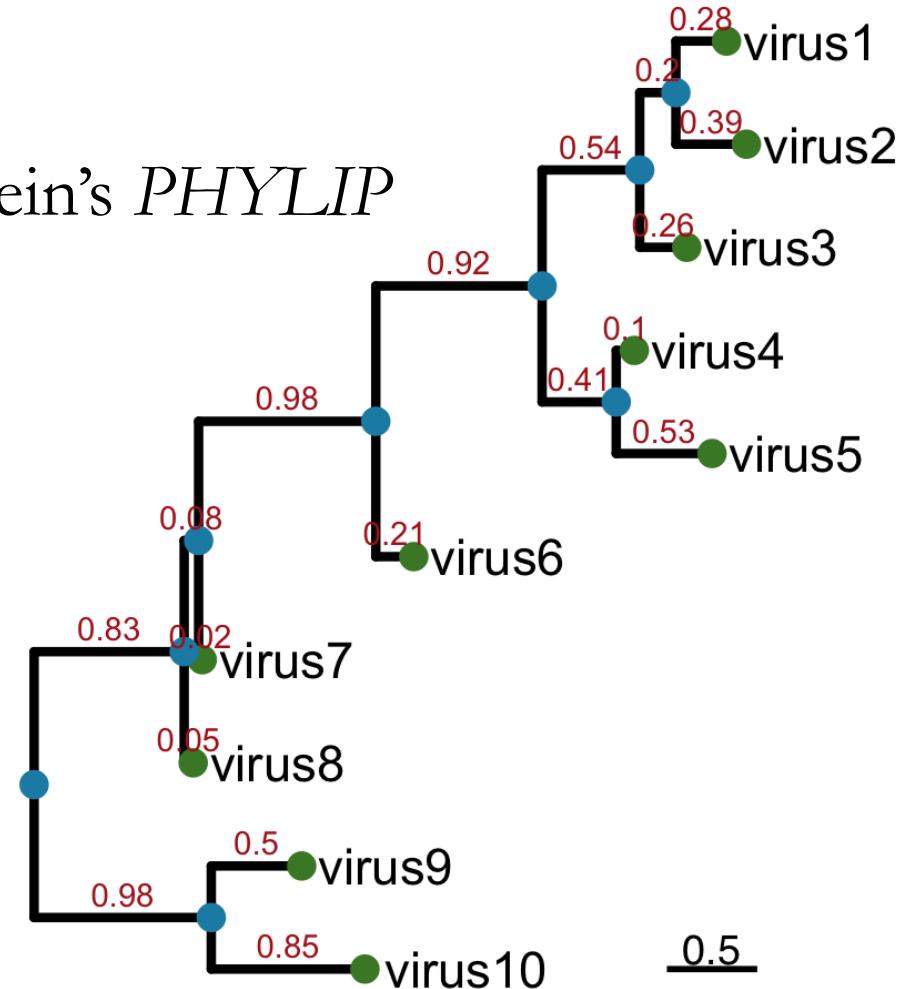


# How to represent phylogenetic tree in computer-readable form?

## Newick format

Developed by Meacham in 1984 in Felsenstein's *PHYLIP*

```
((virus8:0.05,((virus6:0.21,  
(((virus2:0.39,virus1):0.28):0.2,  
virus3):0.26):0.54,  
(virus5:0.53,virus4))):0.1):0.41):0.92):0.98,  
virus7)):0.02):0.08):0.83,  
((virus10:0.85,virus9)):0.5):0.98);
```



# How to represent phylogenetic tree in computer-readable form?

## NEXUS format

Maddison DR *etc.* NEXUS: An extensible file format for systematic information, *Systematic Biology*, 1997.

- Organized into major units known as blocks.
- Some blocks are recognized by most of the programs
- Other blocks are private blocks (recognized by only one program)

### Basic blocks [edit]

#### TAXA block

The TAXA block contains information about taxa.

#### DATA block

The DATA block contains the data matrix (e.g. sequence alignment).

#### TREES block

The TREES block contains phylogenetic trees described using the [Newick format](#), e.g. `((A,B),C);`

```
#NEXUS
BEGIN TAXA;
TAXLABELS A B C;
END;

BEGIN TREES;
TREE tree1 = ((A,B),C);
END;
```

# Problems and issues

- Tree structure is not human-friendly
  - Newick
  - Nexus
  - Many other formats that are not compatible with each other
- Heterogeneous associated data
- Annotating tree

# Problems and issues

- Tree structure is not human-friendly
- **Heterogeneous associated data**
  - Phenotypic data
  - Experimental data
  - Clinical data
  - Analysis findings
- Annotating tree

# Problems and issues

- Tree structure is not human-friendly
- Heterogeneous associated data
- **Annotating tree**
  - *FigTree, TreeDyn, iTOL, EvolView etc.*
  - Restricted to pre-defined annotating utilities
  - Focus on specific analyses and domain specific data
  - Non-overlapping features and incompatibility among tools
  - hard to customize

## Heterogeneous data integration and link to phylogeny

- *treeio* for parsing tree with associated data and mapping external data to phylogeny

## Phylogeny visualization with heterogeneous associated data

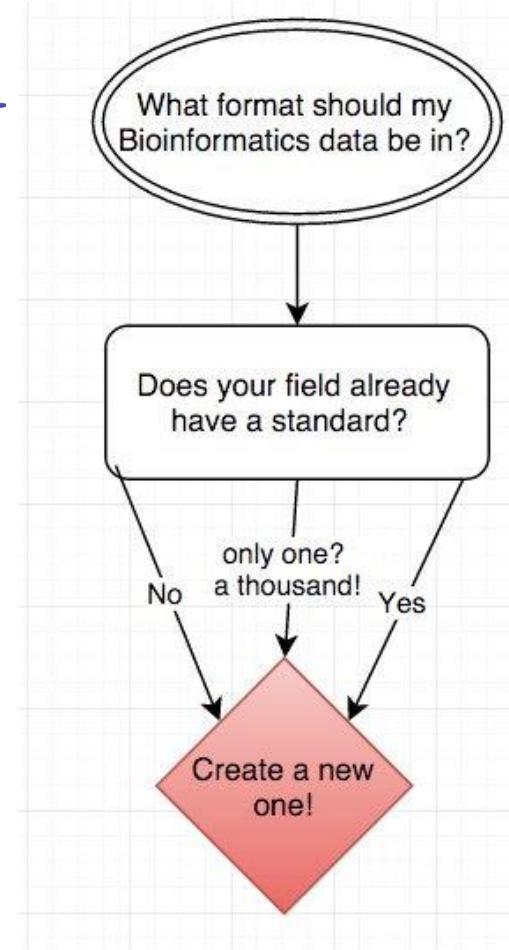
- *ggtree* for visualization and annotation of phylogenetic tree



*treeio*: a platform for tree data integration

# Getting trees into *R* is limited

- Newick and NEXUS
  - ape, phylobase, phytools, rncl
- NeXML
  - RNeXML
- SIMMAP output
  - phyext2 and phytools
- BEAST and MrBayes output
  - PHYLOCH



Phylogenetic program outputs (tree + analysis findings) are not supported.  
External data is hard to be mapped to the structure of phylogeny

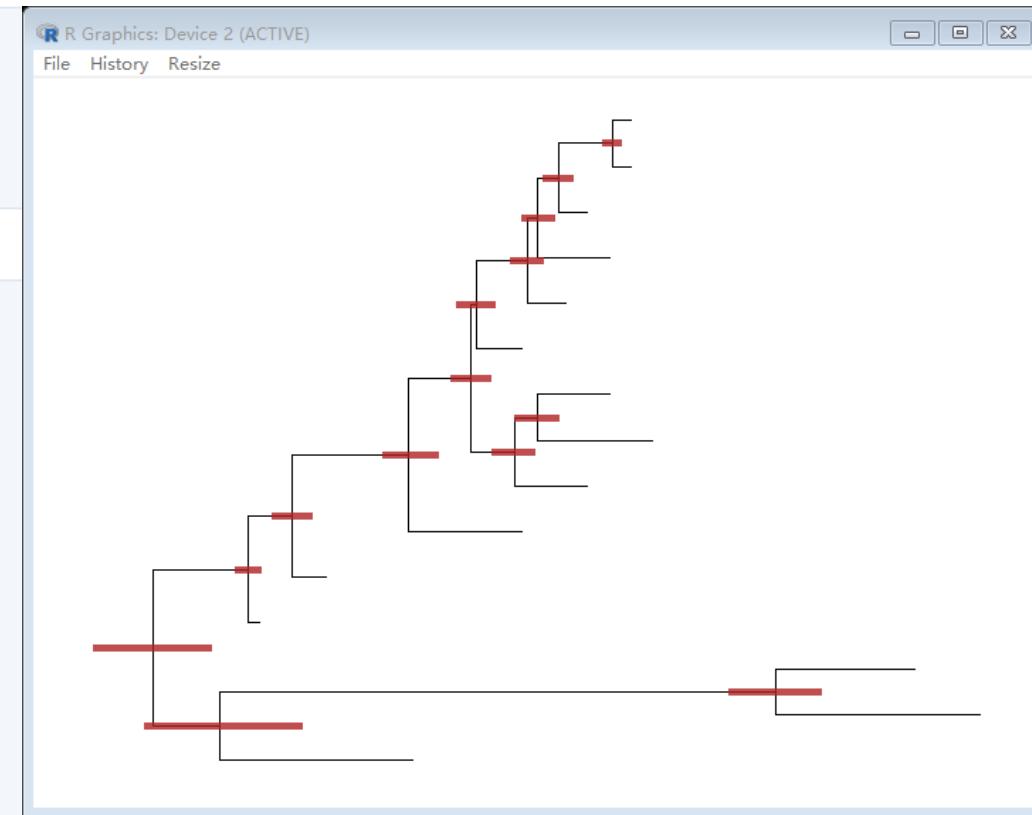
# *treeio*: a platform for tree data integration

Standard formats	Software outputs
Newick	BEAST/MrBayes (NEXUS + metacommments)
NEXUS	HYPHY
New Hampshire eXtended format (NHX)	IQ-TREE (Newick, SH_aLRT/UFboot as node labels)
phylip (phylip msa + Newick)	PAML
jplace (based on json)	PHYLODOG (NHX)
	pplacer/EPA (jplace)
	R8s
	RAxML (Newick + bootstrap as metacomment)
	RevBayes (NHX)

# Importing richly annotated data from software outputs

```
file <- system.file("extdata/BEAST", "beast_mcc.tree", package="treeio")
beast <- read.beast(file)
beast

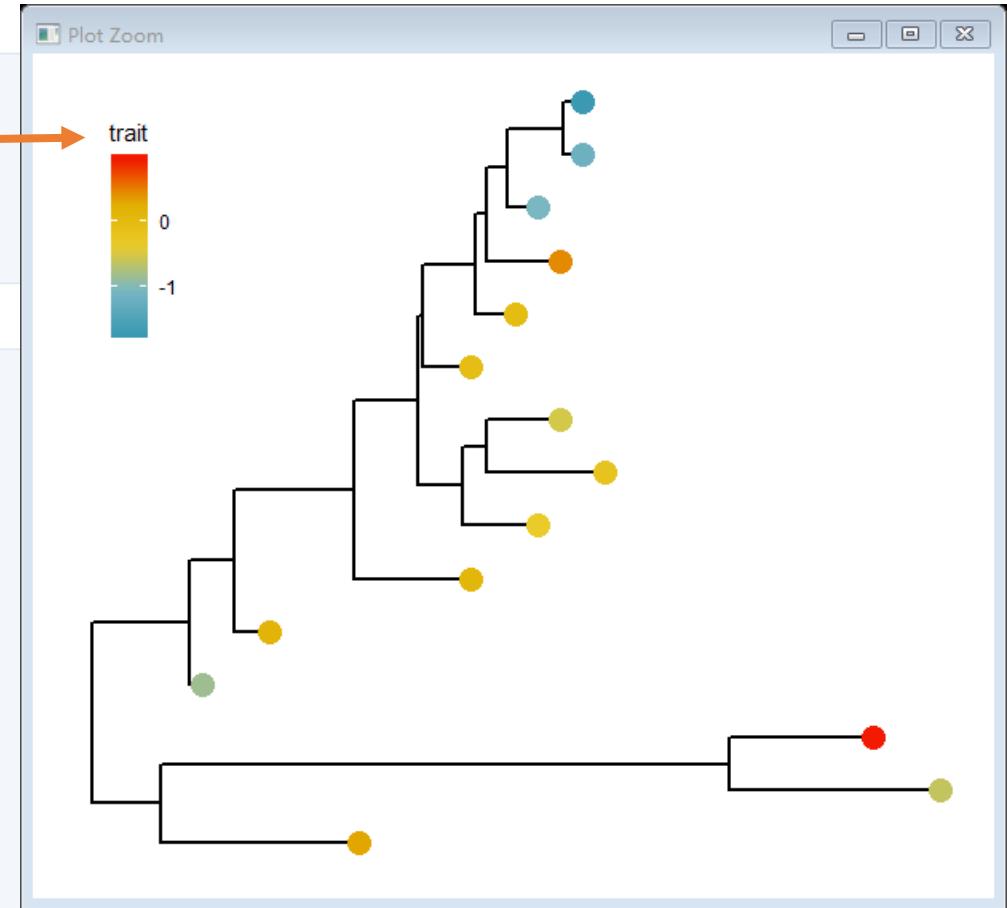
## 'treedata' S4 object that stored information of
## '/tmp/RtmpS6FeFf/Rinst6e0d68e74247/treeio/extdata/BEAST/beast_mcc.tree'.
##
## ...@ phylo:
## Phylogenetic tree with 15 tips and 14 internal nodes.
##
## Tip labels:
## A_1995, B_1996, C_1995, D_1987, E_1996, F_1997, ...
##
## Rooted; includes branch lengths.
##
## with the following features available:
## 'height', 'height_0.95_HPD', 'height_median', 'height_range', 'length',
## 'length_0.95_HPD', 'length_median', 'length_range', 'posterior', 'rate',
## 'rate_0.95_HPD', 'rate_median', 'rate_range'.
```



# Linking external data to phylogeny

```
d <- data_frame(label = as.phylo(beast)$tip.label,
                 trait = rnorm(Ntip(beast)))
tree <- full_join(beast, d, by = 'label')
tree
```

```
## 'treedata' S4 object that stored information of
## 'C:/Users/YGC/Library/R/library/treeio/extdata/BEAST/beast_mcc.tree'.
##
## ...@ phylo:
## Phylogenetic tree with 15 tips and 14 internal nodes.
##
## Tip labels:
## A_1995, B_1996, C_1995, D_1987, E_1996, F_1997, ...
##
## Rooted; includes branch lengths.
##
## with the following features available:
## 'height', 'height_0.95_HPD', 'height_median', 'height_range', 'length',
## 'length_0.95_HPD', 'length_median', 'length_range', 'posterior', 'rate',
## 'rate_0.95_HPD', 'rate_median', 'rate_range', 'trait'.
```



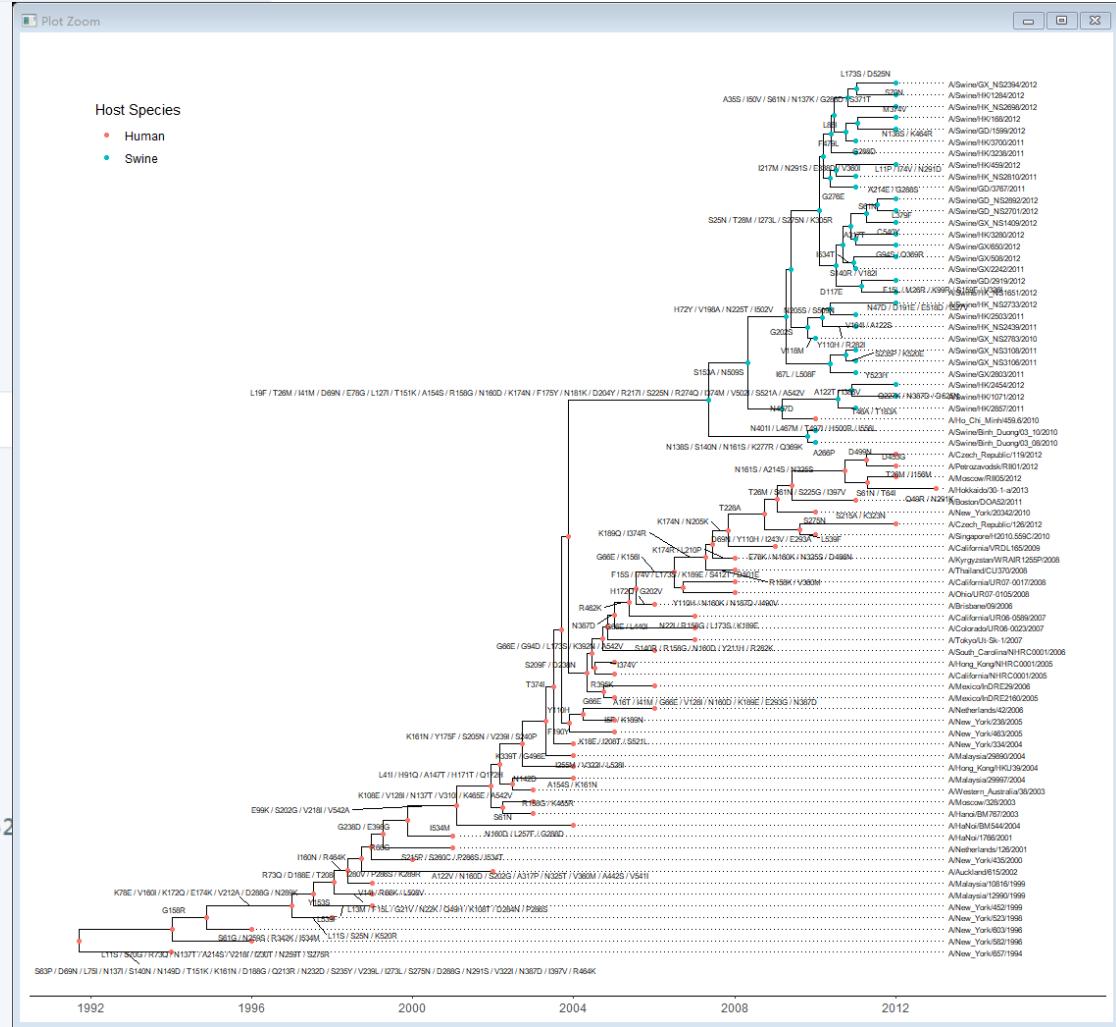
# Combining tree data

```
beast_file <- system.file("examples/MCC_FluA_H3.tree", package="ggtree")
beast_tree <- read.beast(beast_file)

rst_file <- system.file("examples/rst", package="ggtree")
mlc_file <- system.file("examples/mlc", package="ggtree")
codeml_tree <- read.codeml(rst_file, mlc_file)

merged_tree <- merge_tree(beast_tree, codeml_tree)
merged_tree

## 'treedata' S4 object that stored information of
## 'C:/Users/YGC/Library/R/library/ggtree/examples/MCC_FluA_H3.tree',
## 'C:/Users/YGC/Library/R/library/ggtree/examples/rst',
## 'C:/Users/YGC/Library/R/library/ggtree/examples/mlc'.
##
## ...@ phylo:
## Phylogenetic tree with 76 tips and 75 internal nodes.
##
## Tip labels:
## A/Hokkaido/30-1-a/2013, A/New_York/334/2004, A/New_York/463/2005, A/New_York/452
## 
## Rooted; includes branch lengths.
##
## with the following features available:
## 'height', 'height_0.95_HPD', 'height_median', 'height_range', 'length',
## 'length_0.95_HPD', 'length_median', 'length_range', 'posterior', 'rate',
## 'rate_0.95_HPD', 'rate_median', 'rate_range', 'subs', 'AA_subs', 't', 'N',
## 'S', 'dN_vs_dS', 'dN', 'dS', 'N_x_dN', 'S_x_dS'.
```



# Phylogenetic tree as tidy data frame

```
library(tidytree)
as_tibble(merged_tree)
```

```
## # A tibble: 151 x 27
##   parent node branch.length label height height_0.95_HPD height_median
##   <int> <int>      <dbl> <chr>    <dbl> <list>           <dbl>
## 1    115     1       1.71 A/Ho~     0 <lgl [1]>          NA
## 2     94     2       0.499 A/Ne~     9 <dbl [2]>          9
## 3    149     3       1.10 A/Ne~     8 <dbl [2]>          8
## 4     81     4       1.45 A/Ne~    14 <dbl [2]>         14
## 5    150     5       0.758 A/Ne~    8 <dbl [2]>          8
## 6     80     6       0.986 A/Ne~   15 <dbl [2]>         15
## 7     85     7       1.01 A/Ne~   13 <dbl [2]>         13
## 8     78     8       1.98 A/Ne~   17 <dbl [2]>         17
## 9     79     9       1.12 A/Ne~   17 <dbl [2]>         17
## 10    77    10       2.27 A/Ne~   19 <dbl [2]>         19
## # ... with 141 more rows, and 20 more variables: height_range <list>,
## #   length <dbl>, length_0.95_HPD <list>, length_median <dbl>,
## #   length_range <list>, posterior <dbl>, rate <dbl>,
## #   rate_0.95_HPD <list>, rate_median <dbl>, rate_range <list>,
## #   subs <chr>, AA_subs <chr>, t <dbl>, N <dbl>, S <dbl>, dN_vs_dS <dbl>,
## #   dN <dbl>, dS <dbl>, N_x_dN <dbl>, S_x_dS <dbl>
```

← → ⌂ https://cran.r-project.org/web/packages/tidytree/index.html ☆ 🌐 ○ P | 🎙 :

## tidytree: A Tidy Tool for Phylogenetic Tree Data Manipulation

Phylogenetic tree generally contains multiple components including node, edge, branch and associated data. 'tidytree' provides an approach to convert tree object to tidy data frame as well as provides tidy interfaces to manipulate tree data.

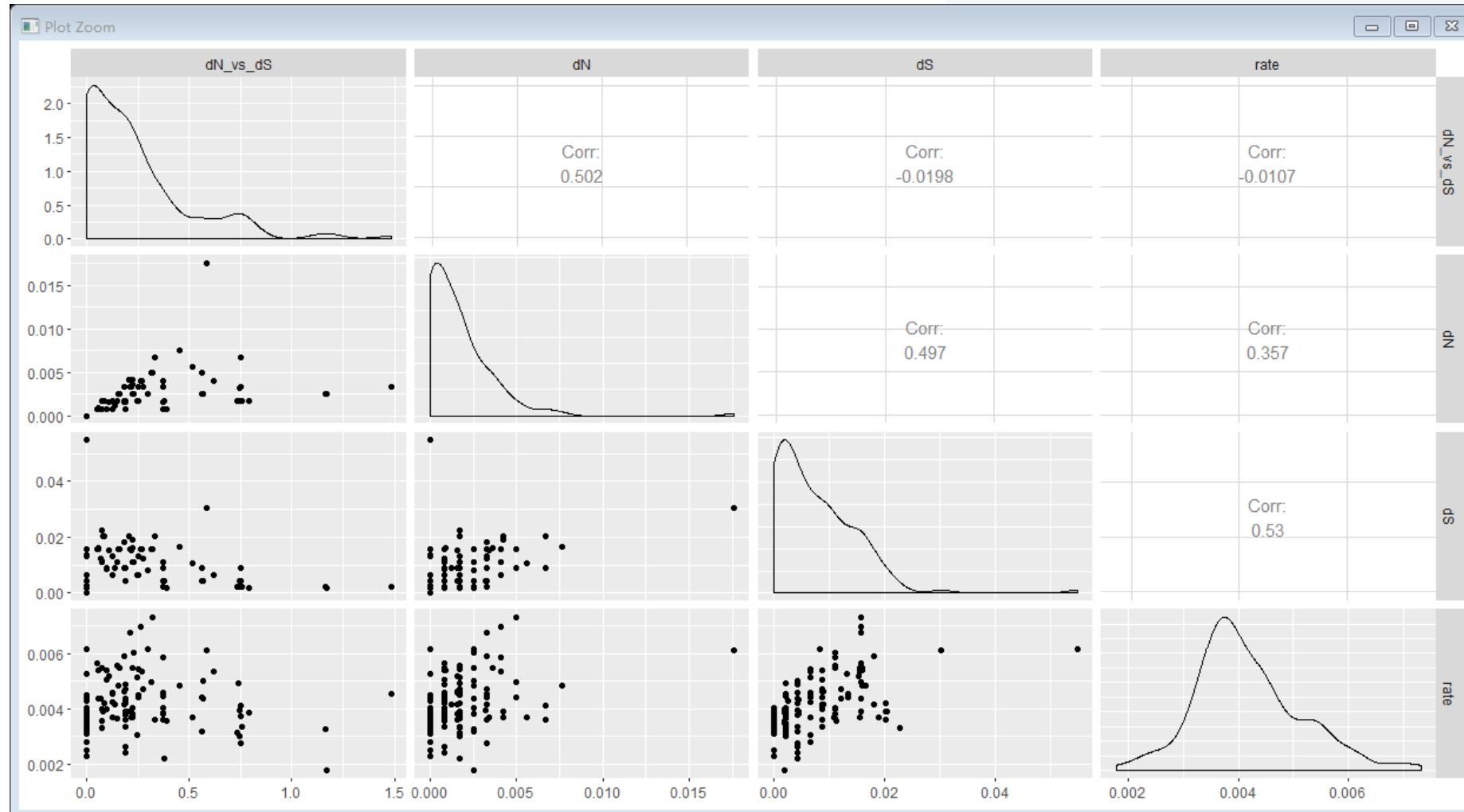
Version: 0.2.1  
Depends: R (≥ 3.4.0)  
Imports: ape, dplyr, lazyeval, magrittr, methods, rlang, tibble, utils  
Suggests: ggtree, knitr, prettydoc, testthat, treeio  
Published: 2018-12-19  
Author: Guangchuang Yu  [aut, cre, cph], Bradley Jones [ctb], Zebulun Arendsee [ctb]  
Maintainer: Guangchuang Yu <guangchuangyu at gmail.com>

```
as_tibble(merged_tree) %>% as.data.frame %>% head(., 2)
```

```
## # A tibble: 2 x 27
##   parent node branch.length label height height_0.95_HPD height_median
##   <int> <int>      <dbl> <chr>    <dbl> <list>           <dbl>
## 1    115     1       1.7132967 A/Hokkaido/30-1-a/2013     0          NA
## 2     94     2       0.4987642 A/New_York/334/2004      9          9, 9
## #   height_median height_range length length_0.95_HPD length_median
## #   <dbl> <list> <dbl> <dbl> <dbl>
## 1        NA 1.7488457  1.241925, 2.352535  1.7134859
## 2         9  9, 9 0.4142844 0.1108738, 0.7635031  0.3910469
## #   length_range posterior rate rate_0.95_HPD
## #   <list> <dbl> <dbl> <dbl>
## 1 1.084697, 3.351183  NA 0.003922044 0.001984250, 0.006131794
## 2 0.03354564, 1.29678831  NA 0.005388027 0.001933106, 0.010083737
## #   rate_median rate_range
## #   <dbl> <list>
## 1 0.003798885 0.001165606, 0.011632788
## 2 0.004924431 0.0009810413, 0.0249303086
## #
## #   subs
## #   <list>
## 1 C72T / A146G / C285T / C369A / G501A / T606C / T873G / C885A / C1242A / A1506G / A1512G
## 2                               T210C / T569A / C720A / C790T / G948A
## #   AA_subs   t     N     S dN_vs_dS   dN     ds N_x_dN S_x_dS
## #   <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Q49R / N291K 0.021 1220.7 465.3  0.0816 0.0017 0.0205    2   9.5
## 2 F190Y 0.009 1220.7 465.3  0.0936 0.0008 0.0089    1   4.1
```

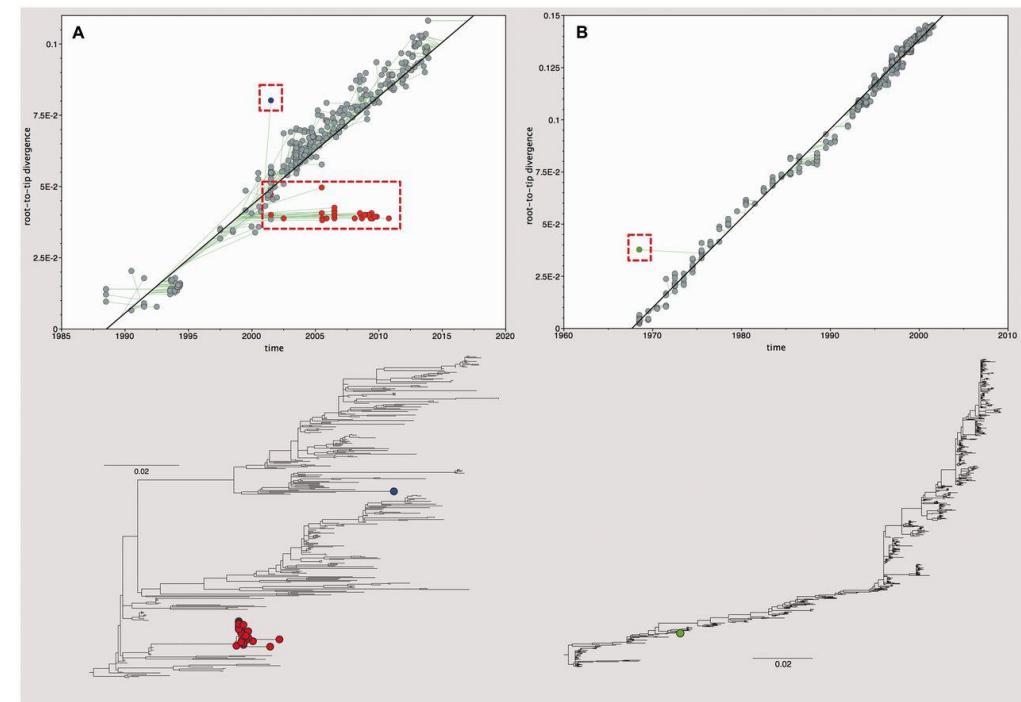
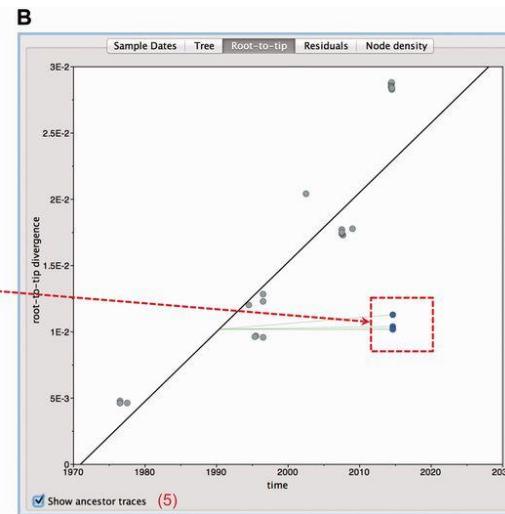
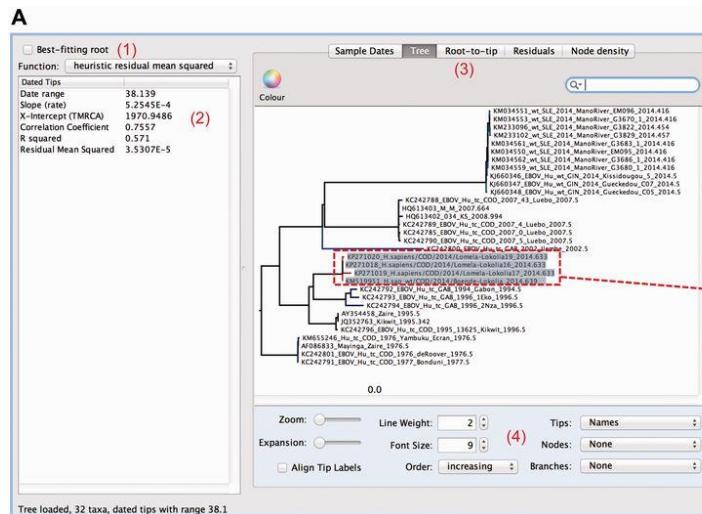
# Data Comparison

```
library(tidytree)
library(ggplot2)
as_tibble(merged_tree) %>%
  dplyr::select(dN_vs_dS, dN, dS, rate) %>%
  subset(dN_vs_dS >=0 & dN_vs_dS <= 1.5) %>%
  GGally::ggpairs(.)
```



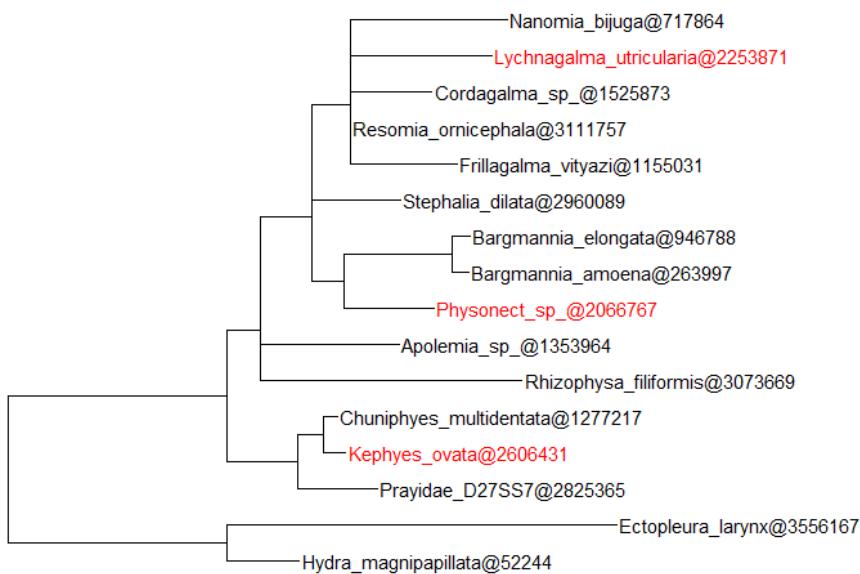
# Removing tips in phylogenetic tree

- Low sequencing quality
- Errors in sequence assembly
- An alignment error in part of the sequence
- An error in phylogenetic inference

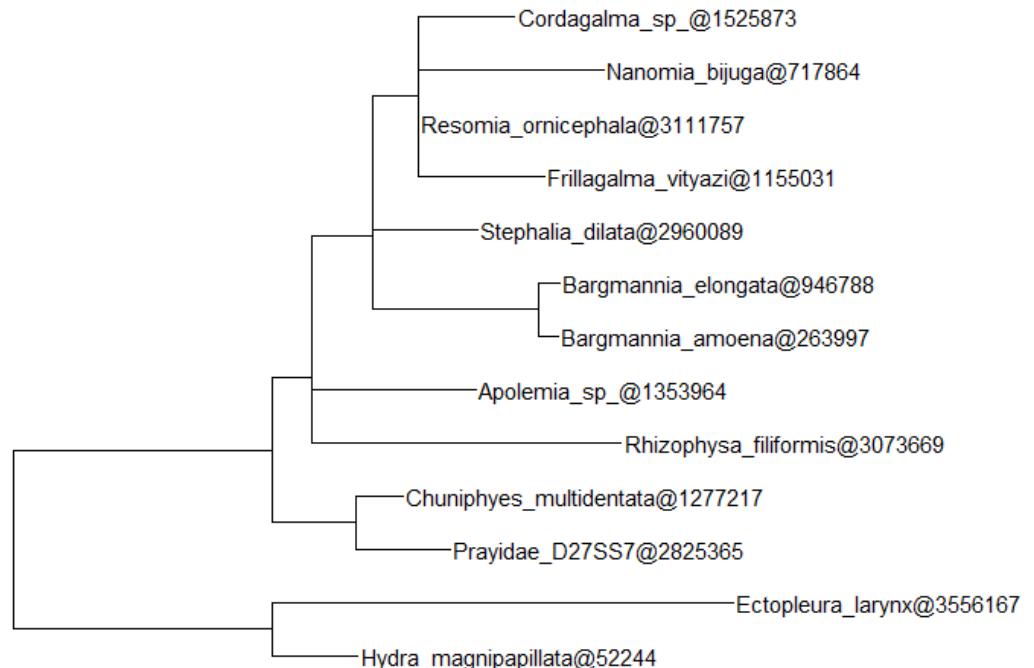


# Removing tips in phylogenetic tree

```
f <- system.file("extdata/NHX", "phyldog.nhx", package="treeio")
nhx <- read.nhx(f)
to_drop <- c("Physonect_sp_@2066767",
           "Lychnagalma_utricularia@2253871",
           "Kephyes_ovata@2606431")
ggtree(nhx) + geom_tiplab(aes(color = label %in% to_drop)) +
  scale_color_manual(values=c('black', 'red')) + xlim(0, 0.7)
```

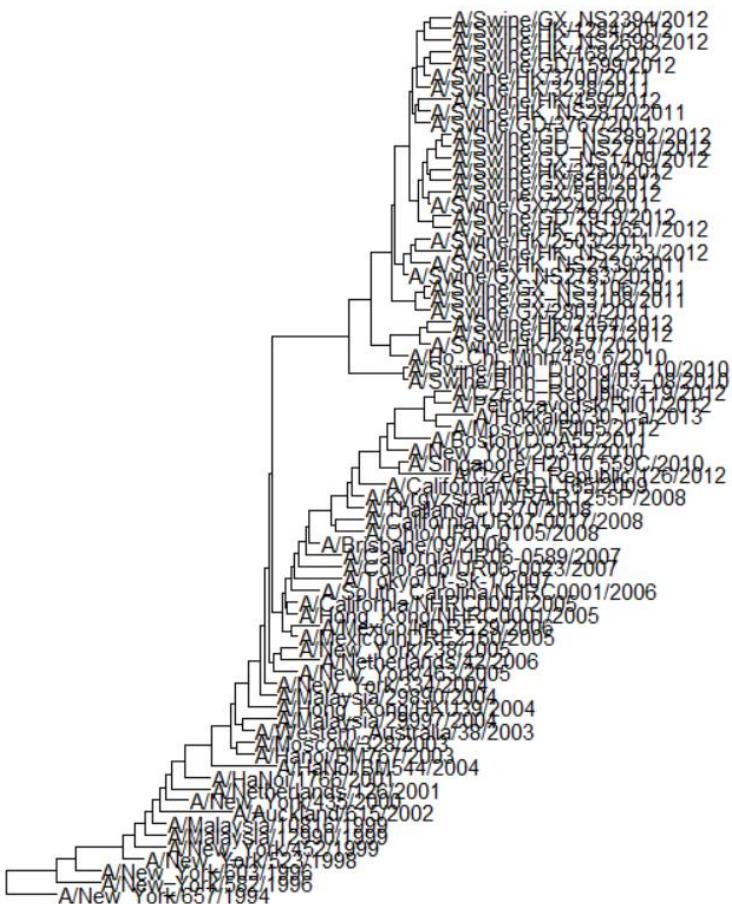


```
nhx_reduced <- drop.tip(nhx, to_drop)
ggtree(nhx_reduced) + geom_tiplab() + xlim(0, 0.7)
```



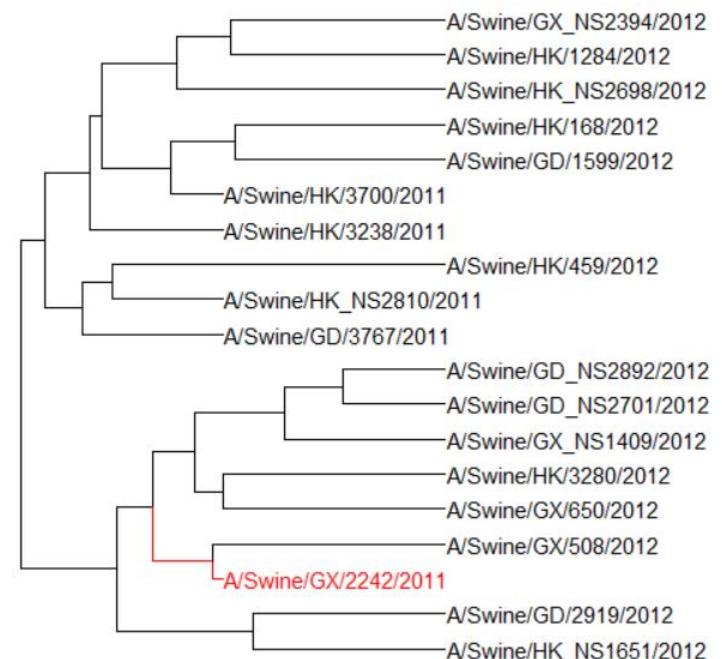
# Subsetting phylogenetic tree

```
ggtree(merged_tree) + geom_tiplab() + xlim(0, 40)
```



```
merged_subset <- tree_subset(merged_tree,  
"A/Swine/GX/2242/2011",  
levels_back = 4)
```

```
ggtree(merged_subset,aes(color=group)) +  
geom_tiplab() +  
xlim(0, 4) +  
scale_color_manual(values=c('black', 'red'))
```



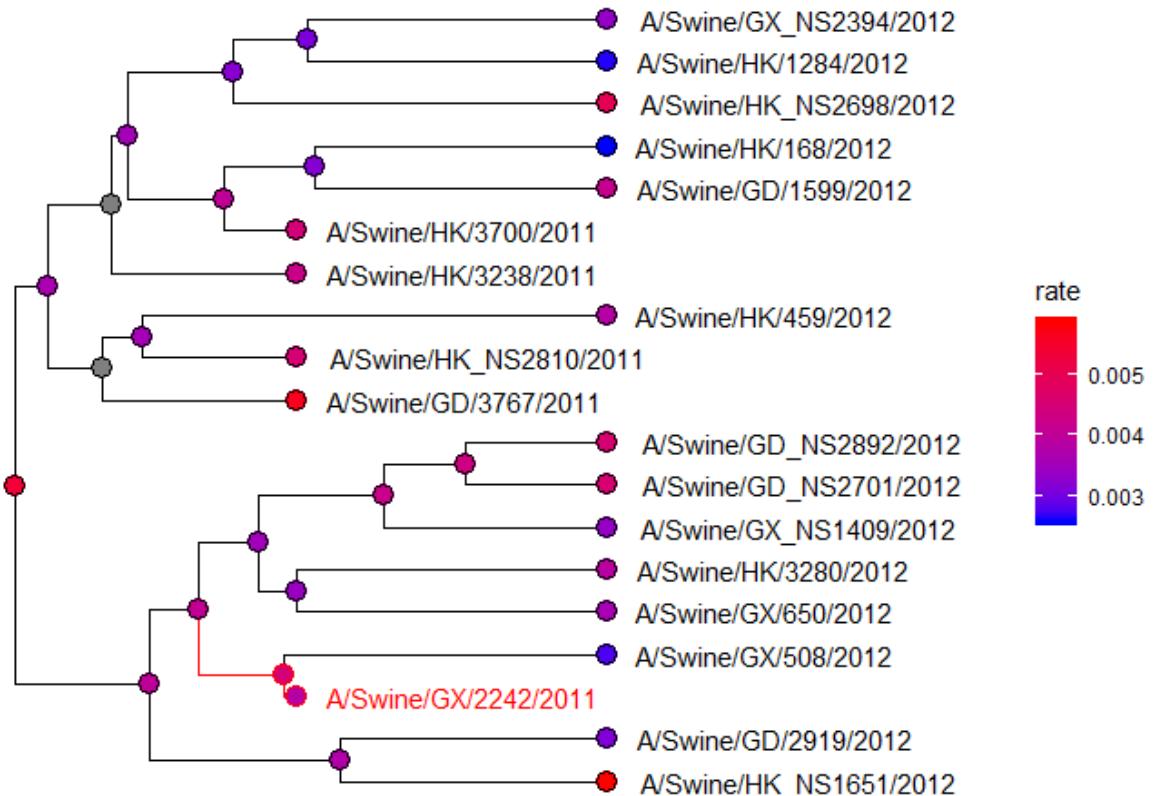
# Subsetting phylogenetic tree

```
merged_subset <- tree_subset(merged_tree,
                             "A/Swine/GX/2242/2011",
                             levels_back = 4)
```

```
## 'treedata' S4 object that stored information of
## 'C:/Users/YGC/Library/R/library/ggtree/examples/MCC_FluA_H3.tree',
## 'C:/Users/YGC/Library/R/library/ggtree/examples/rst',
## 'C:/Users/YGC/Library/R/library/ggtree/examples/mlc'.
##
## ...@ phylo:
## Phylogenetic tree with 19 tips and 18 internal nodes.
##
## Tip labels:
## A/Swine/GD/3767/2011, A/Swine/GD/1599/2012, A/Swine/GX/2242/2011, A/Swine/GX/508/2012, A/Swine
## 
## Rooted; includes branch lengths.
##
## with the following features available:
## 'height', 'height_0.95_HPD', 'height_median', 'height_range', 'length',
## 'length_0.95_HPD', 'length_median', 'length_range', 'posterior', 'rate',
## 'rate_0.95_HPD', 'rate_median', 'rate_range', 'subs', 'AA_subs', 't', 'N',
## 'S', 'dN_vs_dS', 'dN', 'dS', 'N_x_dN', 'S_x_dS'.
```

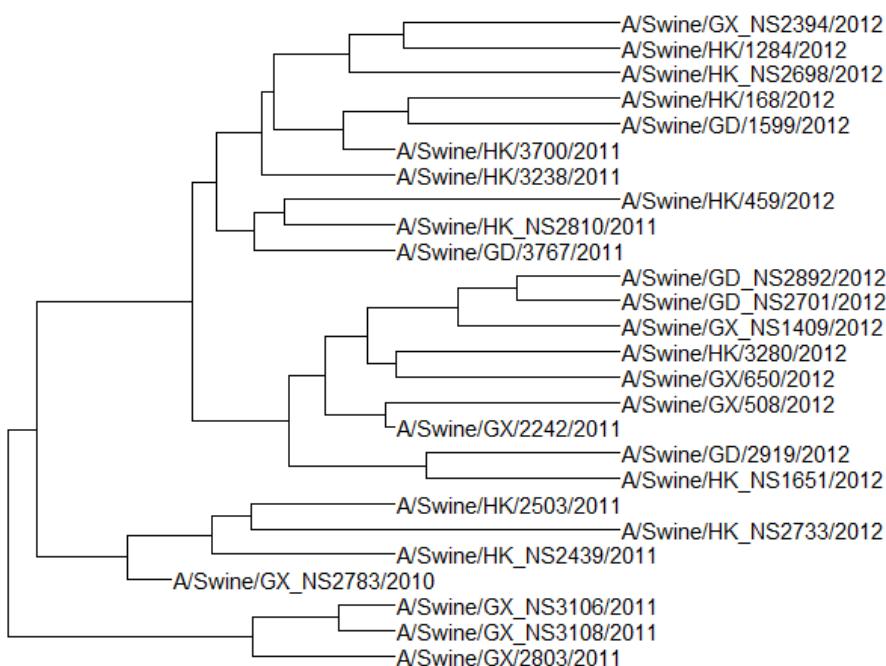
  

```
ggtree(merged_subset,aes(color=group)) +
  geom_tiplab(hjust = -.1) +
  xlim(0, 3) +
  geom_point(aes(fill=rate), shape=21, size=4) +
  scale_color_manual(values=c('black', 'red'), guide=F) +
  scale_fill_continuous(low='blue', high='red') +
  theme(legend.position = 'right')
```

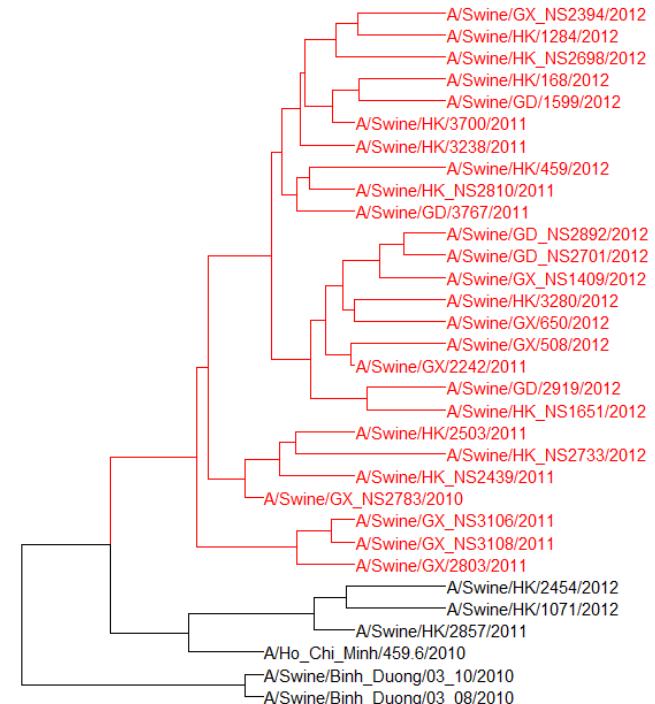


# Subsetting phylogenetic tree

```
clade_subset <- tree_subset(merged_tree, node=121, levels_back=0)
ggtree(clade_subset) +
  geom_tiplab() +
  xlim(0, 4)
```

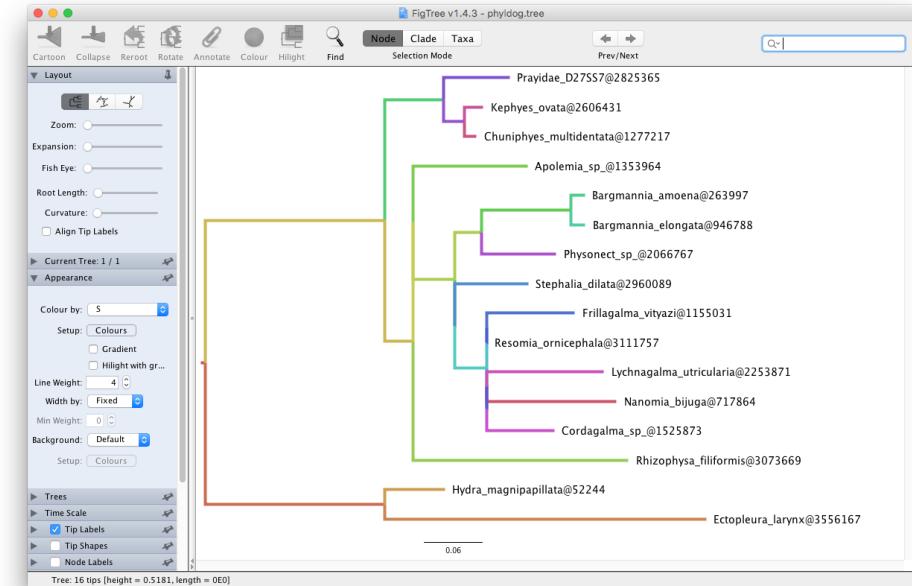
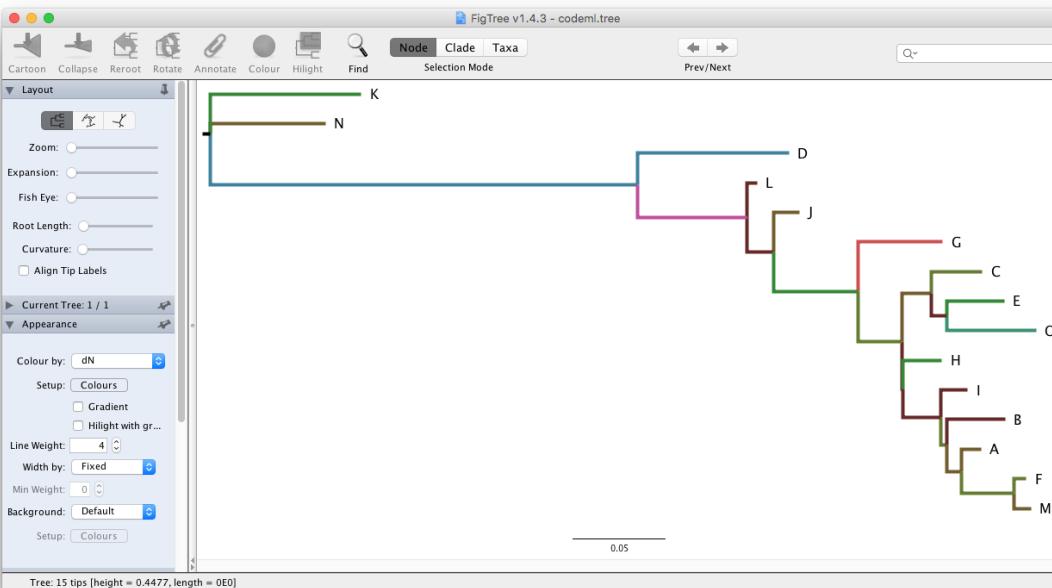


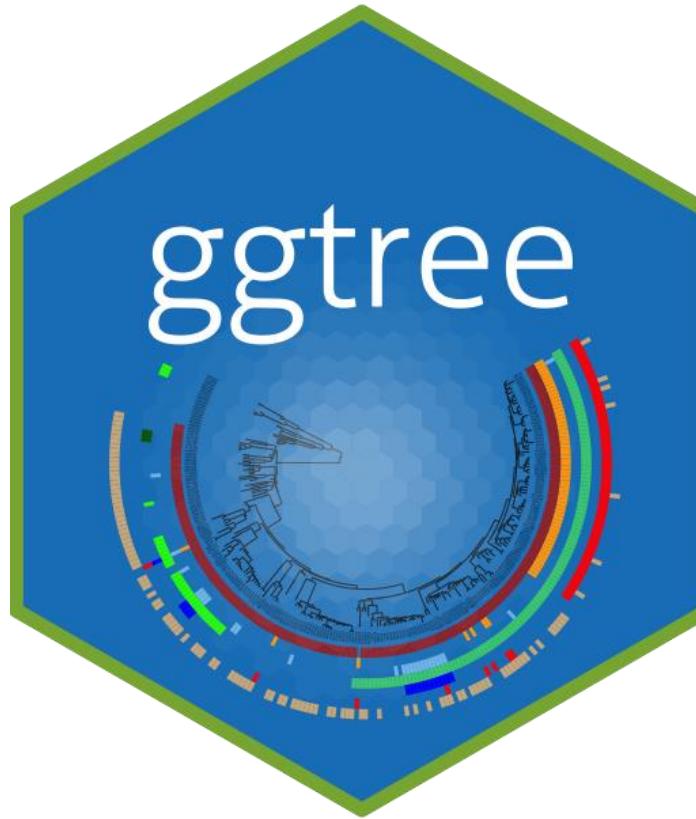
```
clade_subset2 <- tree_subset(merged_tree, node=121, levels_back=2)
ggtree(clade_subset2,aes(color=group)) +
  geom_tiplab() +
  xlim(0, 7.5) +
  scale_color_manual(values=c('black', 'red'))
```



# Exporting trees with data

- Converting software output
- Combining tree with external data
- Merging tree data from different sources





*ggtree*: tree visualization and annotation

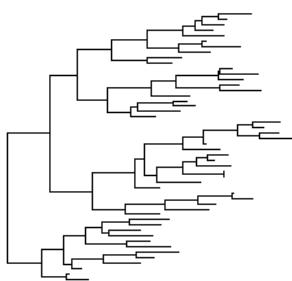
# *ggtree*: tree visualization and annotation

- Supports tree data parsed by *treeio* and other tree objects
- Integration of various types/sources of data
- High levels of customization using grammar of graphics
- Manipulating and exploring tree visually
- Reproducibility, automation & communication

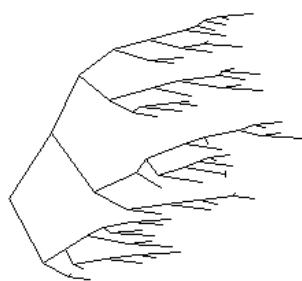
G Yu, D Smith, H Zhu, Y Guan, TTY Lam. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 2017, 8(1):28-36  
doi: 10.1111/2041-210X.12628

# Supported tree layouts

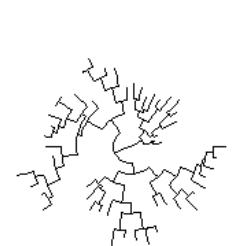
rectangular (phylogram)



slanted (phylogram)



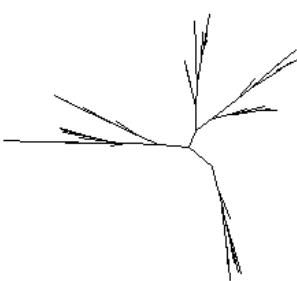
circular (phylogram)



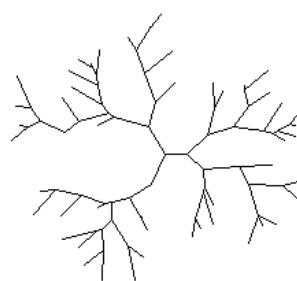
fan (phylogram)



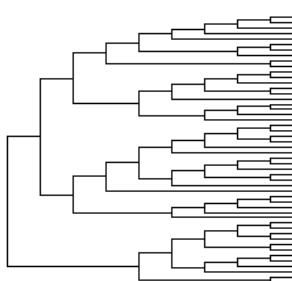
equal angle (unrooted)



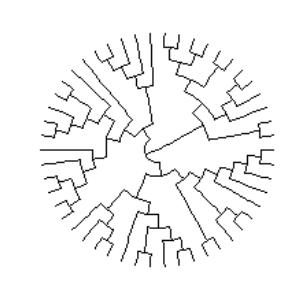
daylight (unrooted)



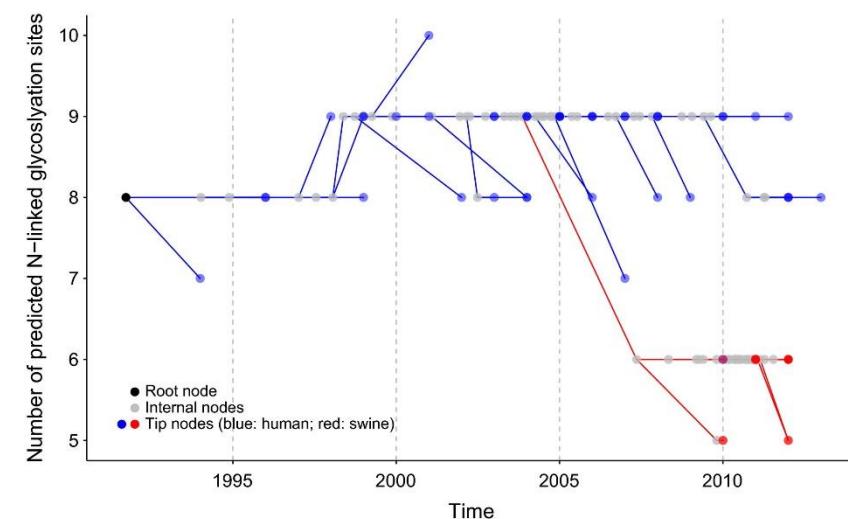
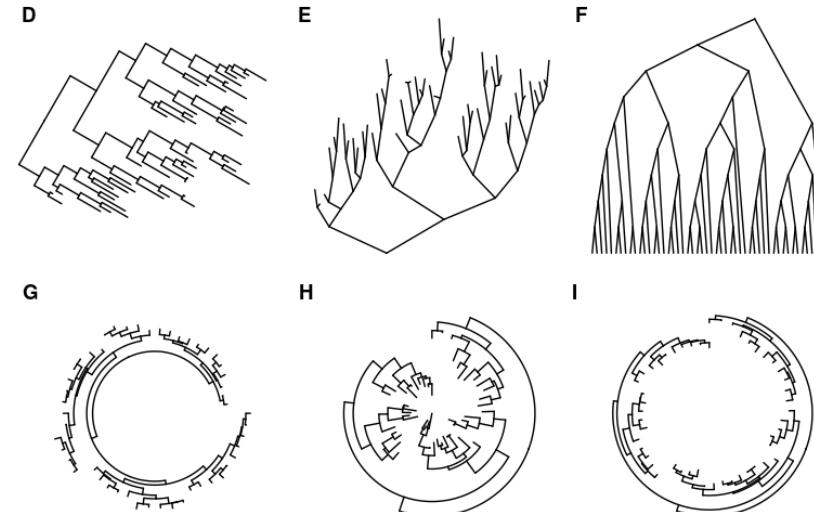
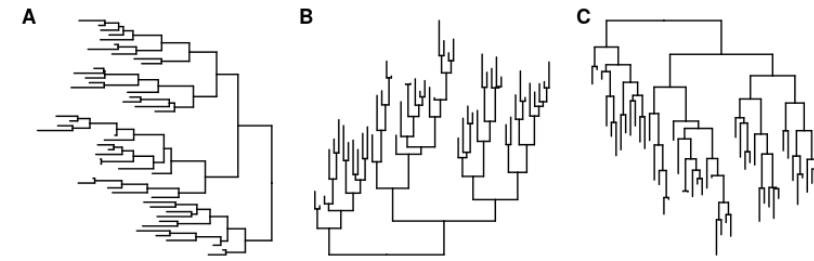
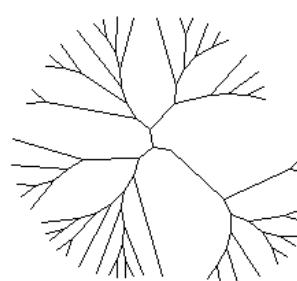
rectangular (cladogram)



circular (cladogram)

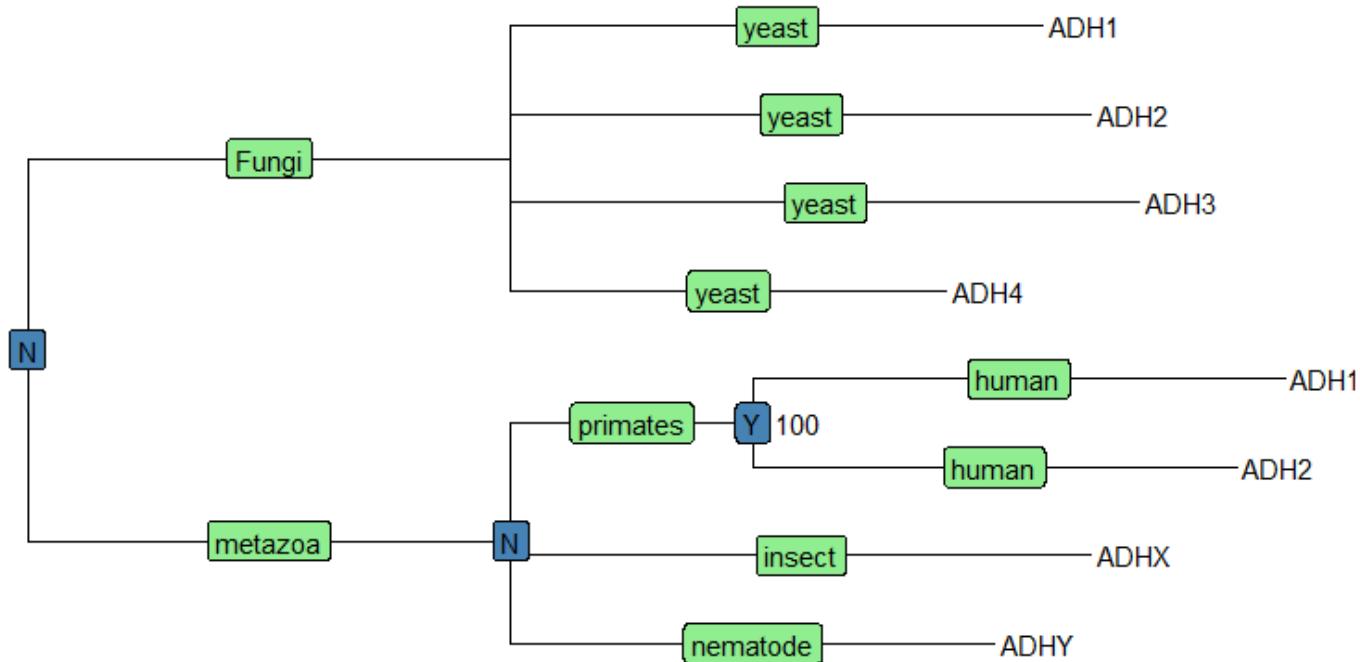


daylight (cladogram)



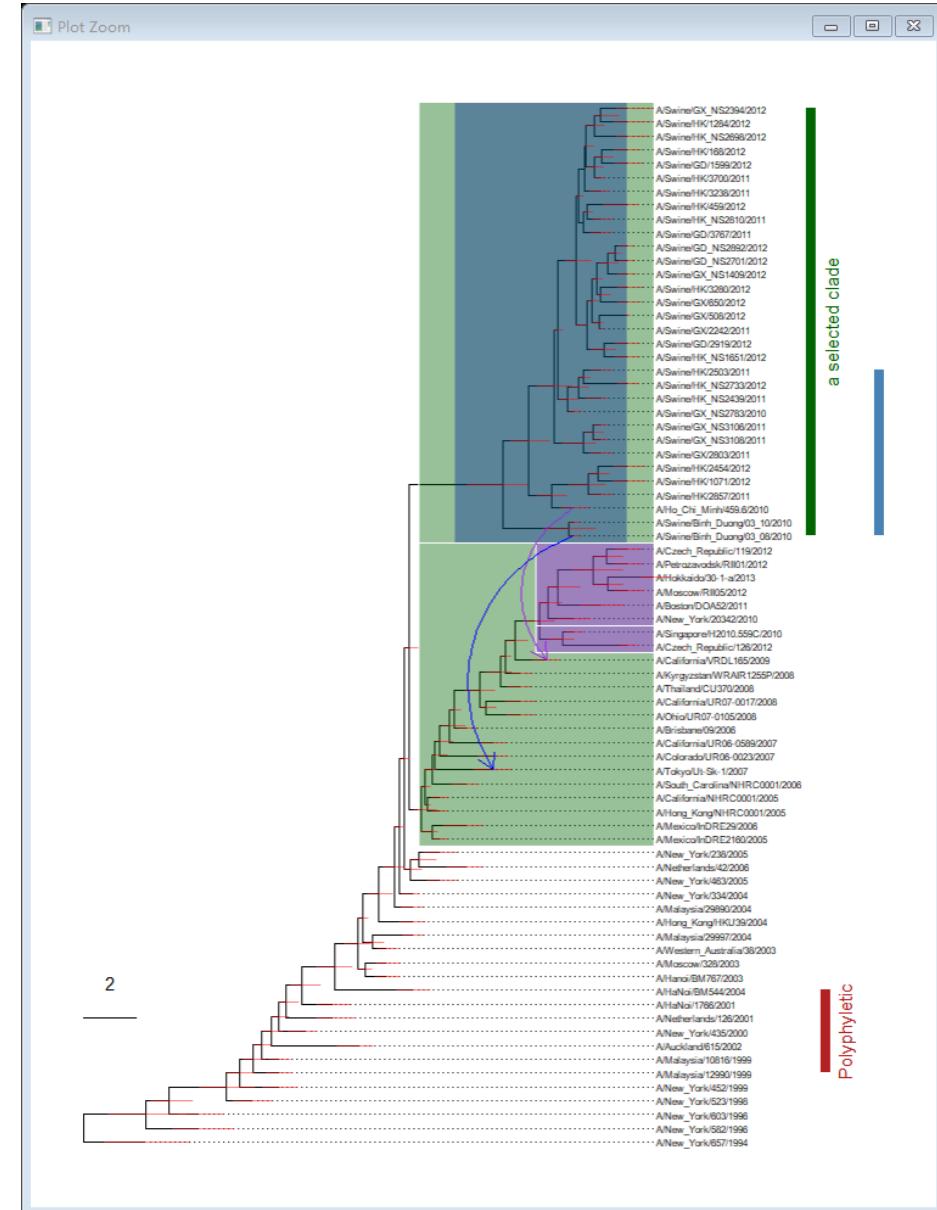
# ggtree: tree visualization using grammar of graphics

```
library(ggtree)
tree <- read.nhx(system.file('extdata/NHX/ADH.nhx', package='treeio'))
ggtree(tree) + geom_tiplab() +
  geom_label(aes(x=branch, label=S), fill='lightgreen') +
  geom_label(aes(label=D), fill='steelblue') +
  geom_text(aes(label=B), hjust=-0.5)
```



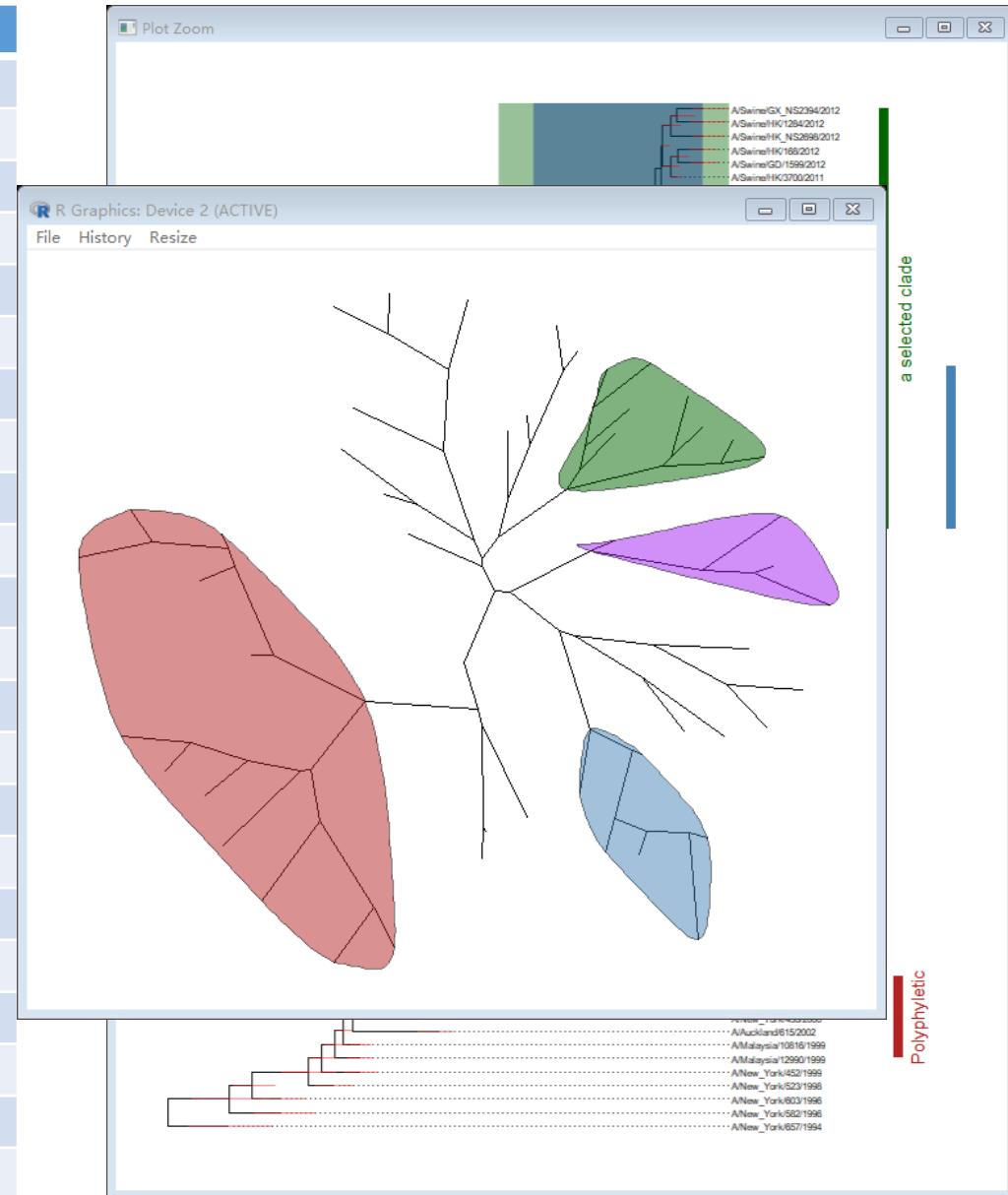
# ggtree: geometric layers for tree annotation

Layer	Description
geom_balance	Highlights the two direct descendant clades of an internal node
geom_cladelabel	Annotate a clade with bar and text label
geom_hilight	Highlight a clade with rectangle
geom_hilight_encircle	Highlight a clade with xspline
geom_label2	Modified version of geom_label, with subsetting supported
geom_motif	Draw aligned motif
geom_nodepoint	Annotate internal nodes with symbolic points
geom_nodelab	Annotate internal nodes with text label
geom_nodelab2	Annotate internal node with text label for circular layout
geom_point2	Modified version of geom_point, with subsetting supported
geom_range	Bar layer to present uncertainty of evolutionary inference
geom_rootpoint	Annotate root node with symbolic point
geom_rootedge	Display root edge
geom_segment2	Modified version of geom_segment, with subsetting supported
geom_strip	Annotate associated taxa with bar and (optional) text label
geom_taxalink	Associate two related taxa by linking them with a curve
geom_text2	Modified version of geom_text, with subsetting supported
geom_tiplab	Layer of tip labels
geom_tiplab2	Layer of tip labels for circular layout
geom_tippoint	Annotate tips with symbolic points
geom_tree	Tree structure layer, with multiple layouts supported
geom_treescale	Tree scale legend



# ggtree: geometric layers for tree annotation

Layer	Description
geom_balance	Highlights the two direct descendant clades of an internal node
geom_cladelabel	Annotate a clade with bar and text label
geom_hilight	Highlight a clade with rectangle
geom_hilight_encircle	Highlight a clade with xspline
geom_label2	Modified version of geom_label, with subsetting supported
geom_motif	Draw aligned motif
geom_nodepoint	Annotate internal nodes with symbolic points
geom_nodelab	Annotate internal nodes with text label
geom_nodelab2	Annotate internal node with text label for circular layout
geom_point2	Modified version of geom_point, with subsetting supported
geom_range	Bar layer to present uncertainty of evolutionary inference
geom_rootpoint	Annotate root node with symbolic point
geom_rootedge	Display root edge
geom_segment2	Modified version of geom_segment, with subsetting supported
geom_strip	Annotate associated taxa with bar and (optional) text label
geom_taxalink	Associate two related taxa by linking them with a curve
geom_text2	Modified version of geom_text, with subsetting supported
geom_tiplab	Layer of tip labels
geom_tiplab2	Layer of tip labels for circular layout
geom_tipoint	Annotate tips with symbolic points
geom_tree	Tree structure layer, with multiple layouts supported
geom_treescale	Tree scale legend



# Overlay organism silhouettes

**PHYLOPIC** [About](#) [Browse](#) [Submit](#) [Dev](#) [Log In](#) [More](#)

Free silhouette images of animals, plants, and other life forms, available for reuse under a [Public Domain](#) or [Creative Commons](#) license.

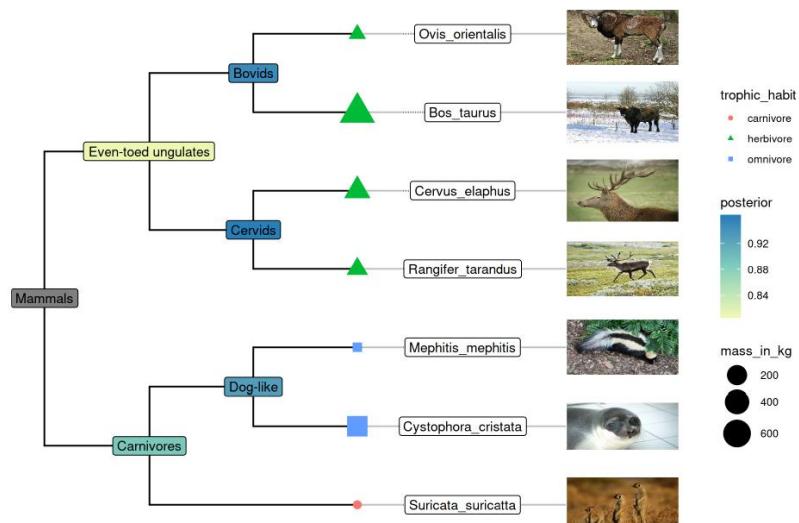
Enter the name of a group of life forms. It may be scientific (e.g., *Homo sapiens*) or informal (e.g., humans).

## Latest Submissions

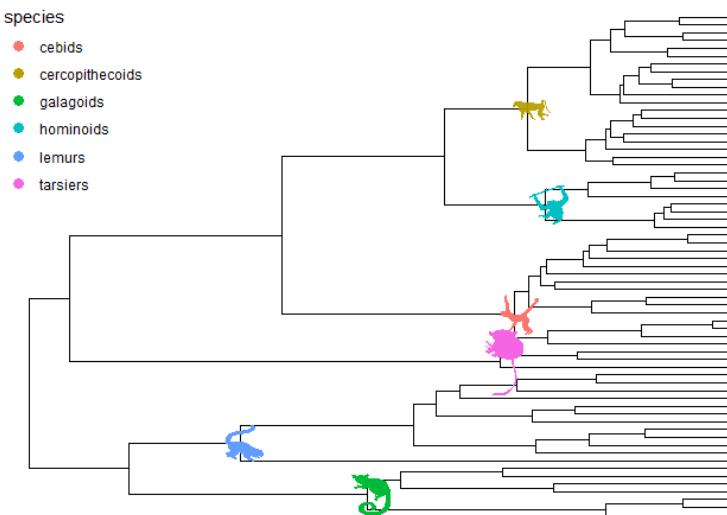


3189 Images in the database. Click on one for more details.

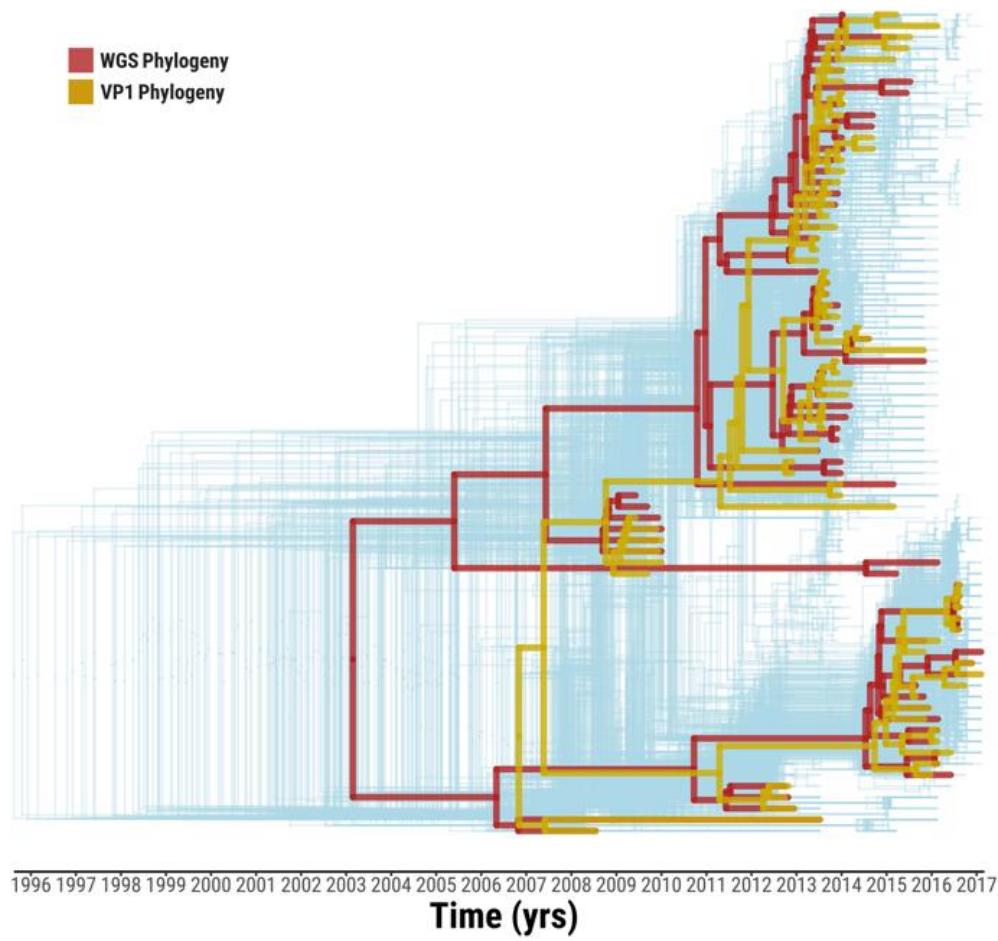


```
tree <- ape::read.nexus("https://raw.githubusercontent.com/rgriff23/Dissertation/master/Chapter_phylopic_info <- data.frame(node = c(124, 113, 110, 96, 89, 70),
phylopic = c("7fb9bea8-e758-4986-afb2-95a2c3bf983d",
"bac25f49-97a4-4aec-beb6-f542158ebd23",
"f598fb39-facf-43ea-a576-1861304b2fe4",
"aceb287d-84cf-46f1-868c-4797c4ac54a8",
"0174801d-15a6-4668-bfe0-4c421fbe51e8",
"72f2f854-f3cd-4666-887c-35d5c256ab0f"),
species = c("galagooids", "lemurs", "tarsiers",
"cebids", "hominoids", "cercopithecoidea"))

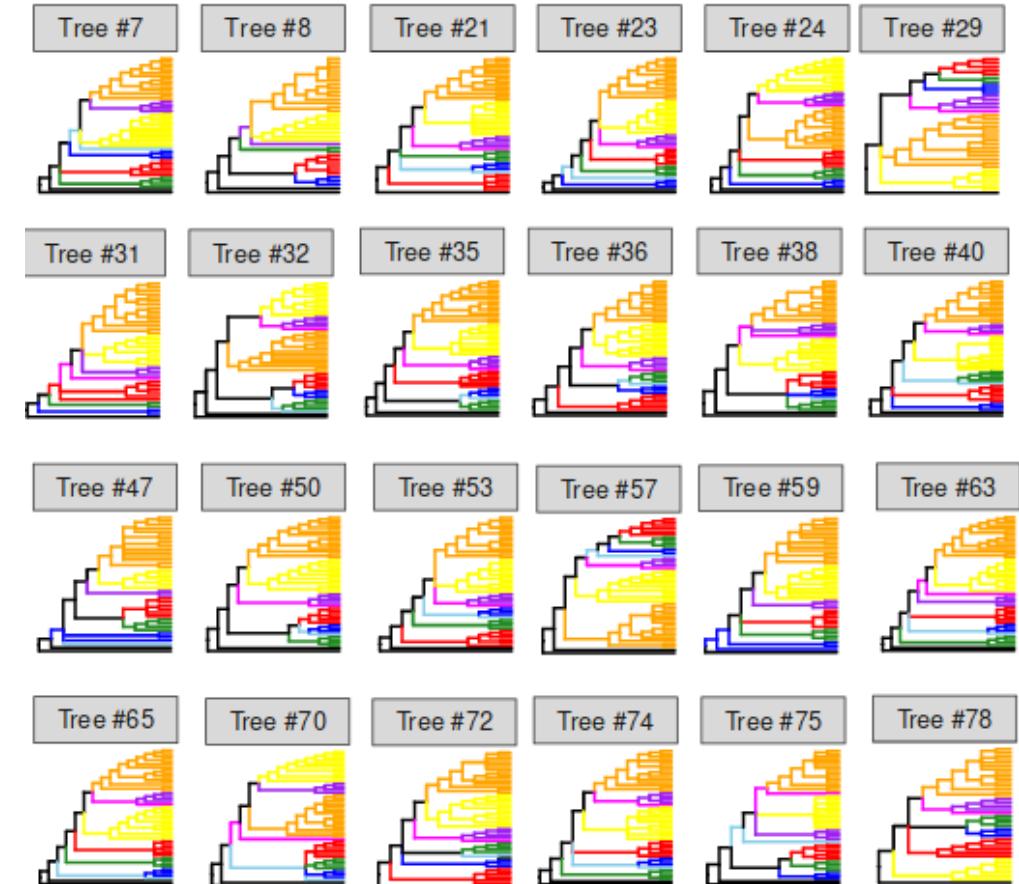
pg <- ggtree(tree)
pg %>%+ phylopic_info +
  geom_nodelab(aes(image=phylopic, color=species, subset=!is.na(species)), geom="phylopic") +
  theme(legend.position=c(.1, .8))
```



# Plot many trees at once



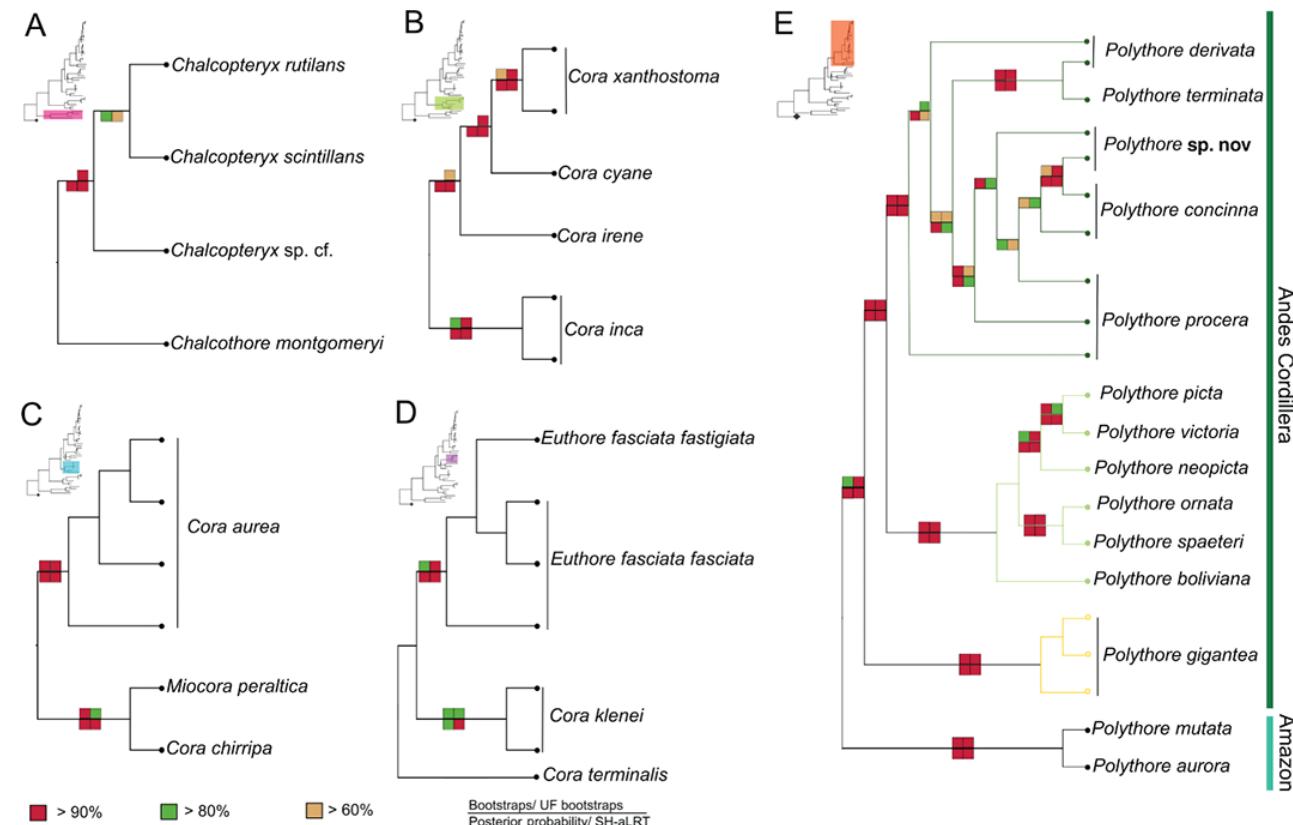
<https://dx.doi.org/10.1038/s41598-018-32693-8>



<https://doi.org/10.1016/j.ympev.2018.07.020>

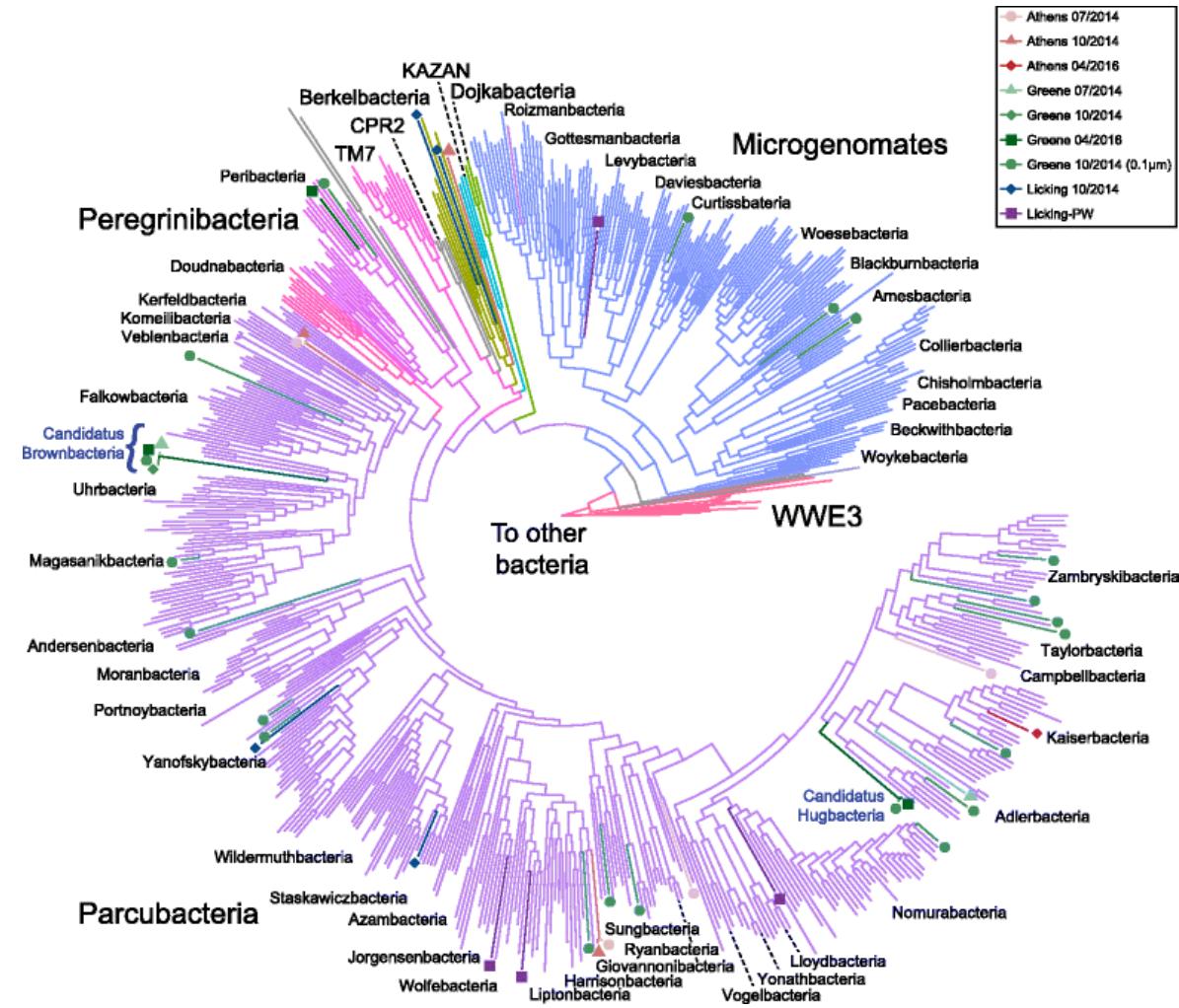
# Tree manipulation

- viewclade
- groupOTU & groupClade
- scaleClade
- collapse & expand
- rotate & flip
- open\_tree & rotate\_tree
- rescale\_tree



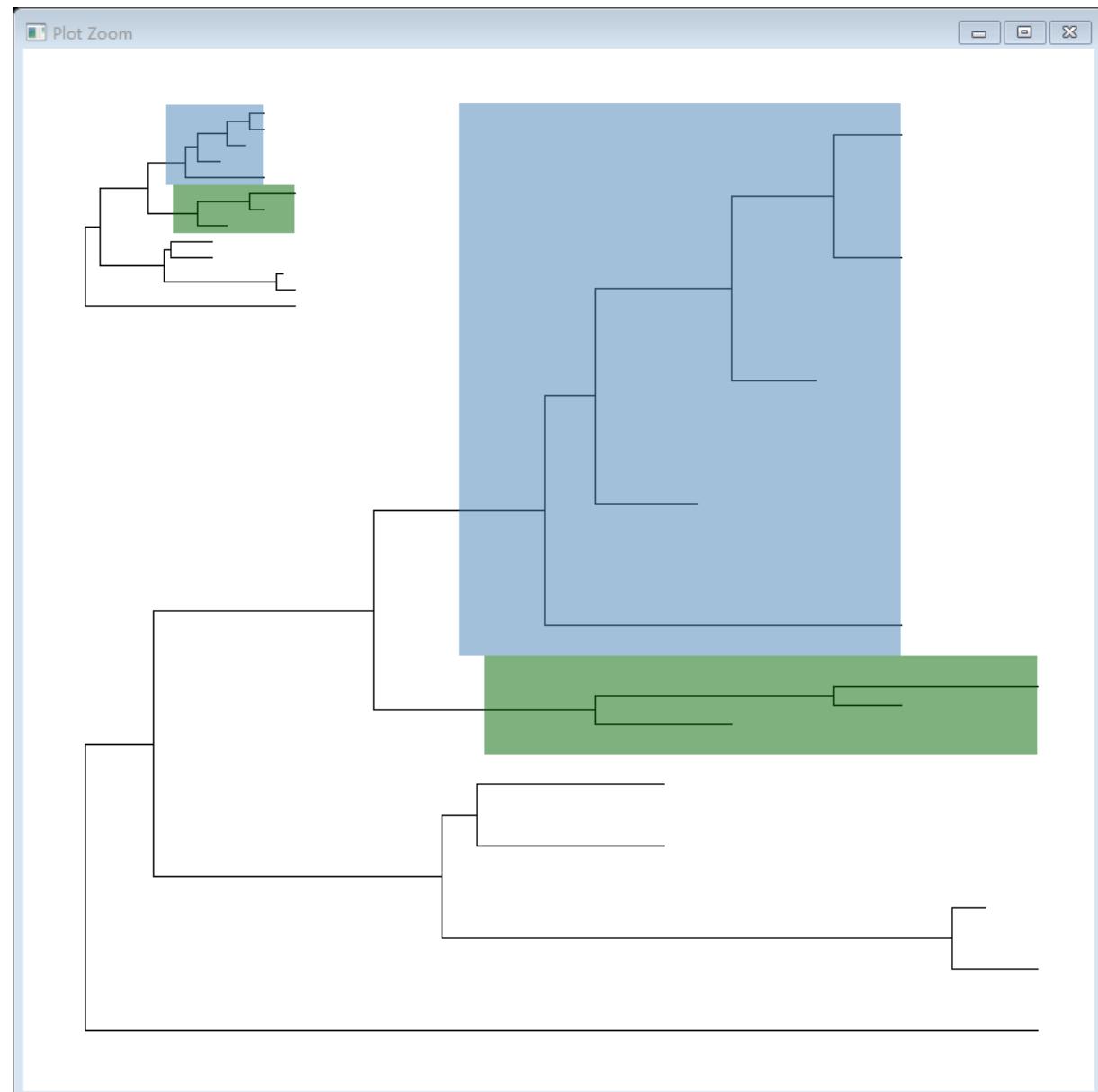
# Tree manipulation

- viewclade
- groupOTU & groupClade
- scaleClade
- collapse & expand
- rotate & flip
- open\_tree & rotate\_tree
- rescale\_tree



# Tree manipulation

- `viewclade`
- `groupOTU` & `groupClade`
- **scaleClade**
- `collapse` & `expand`
- `rotate` & `flip`
- `open_tree` & `rotate_tree`
- `rescale_tree`



# Tree manipulation

- viewclade
- groupOTU & groupClade
- scaleClade
- **collapse & expand**
- rotate & flip
- open\_tree & rotate\_tree
- rescale\_tree

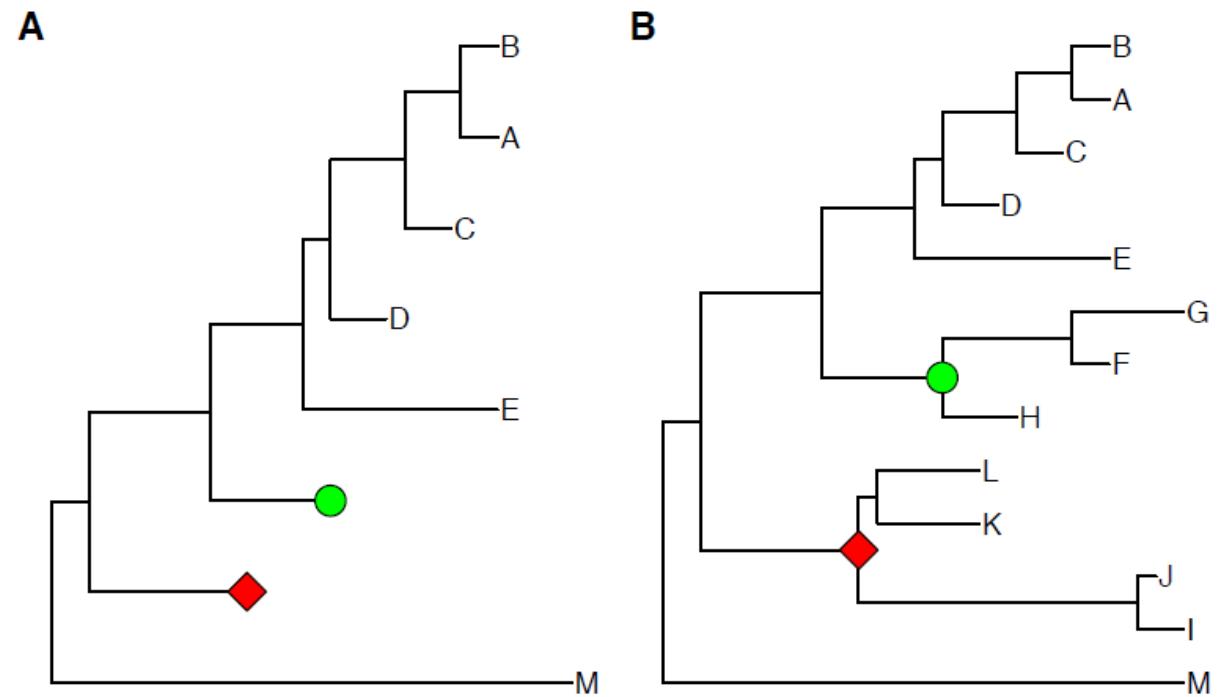


Figure 3.6: **Collapsing selected clades and expanding collapsed clades.** Clades can be selected to collapse (A) and the collapsed clades can be expanded back (B) if necessary as *ggtree* stored all information of species relationships. Green and red symbols were displayed on the tree to indicate the collapsed clades.

# Tree manipulation

- viewclade
- groupOTU & groupClade
- scaleClade
- collapse & expand
- **rotate & flip**
- open\_tree & rotate\_tree
- rescale\_tree

```
p1 <- p + geom_point2(aes(subset=node==16), color='darkgreen', size=5)
p2 <- rotate(p1, 17) %>% rotate(21)
flip(p2, 17, 21)
```

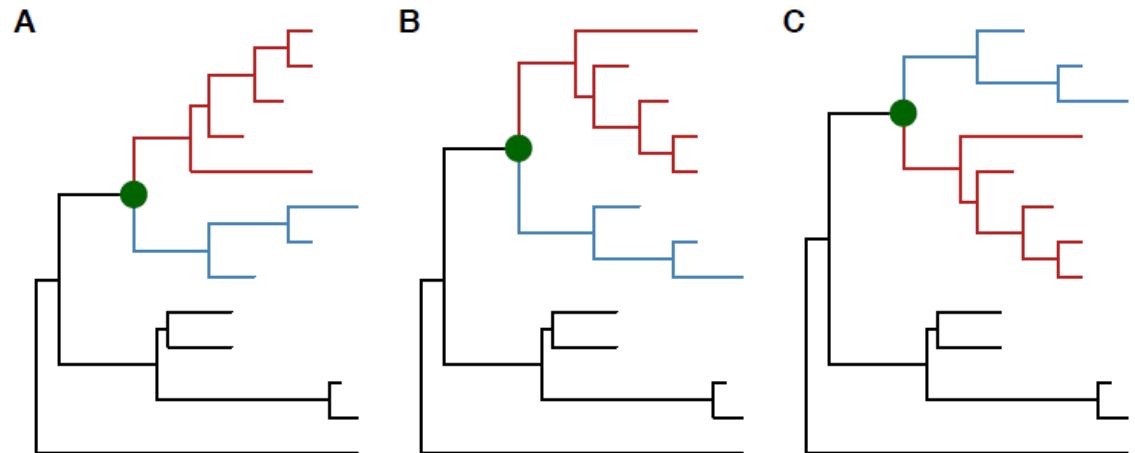


Figure 3.9: Exploring tree structure. A clade (indicated by darkgreen circle) in a tree (A) can be rotated by 180° (B) and the positions of its immediate descendant clades (colored by blue and red) can be exchanged (C).

# Tree manipulation

- viewclade
- groupOTU & groupClade
- scaleClade
- collapse & expand
- rotate & flip
- **open\_tree & rotate\_tree**
- rescale\_tree

```
p3 <- open_tree(p, 180) + geom_tiplab2()  
print(p3)
```

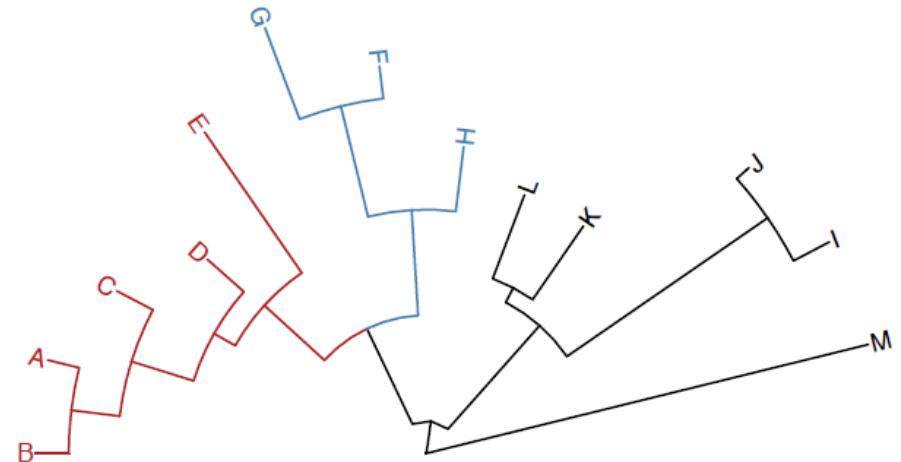


Figure 3.10: **Transforming a tree to fan layout.** A tree can be transformed to fan layout by `open_tree` with specific *angle* parameter.

```
rotate_tree(p3, 180)
```

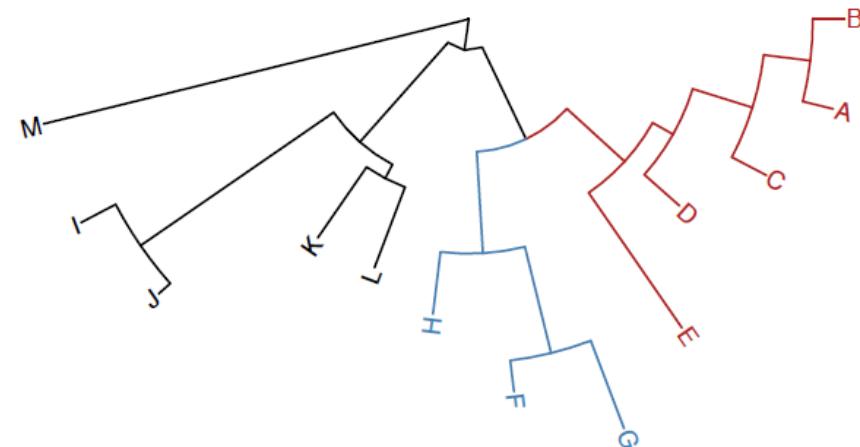
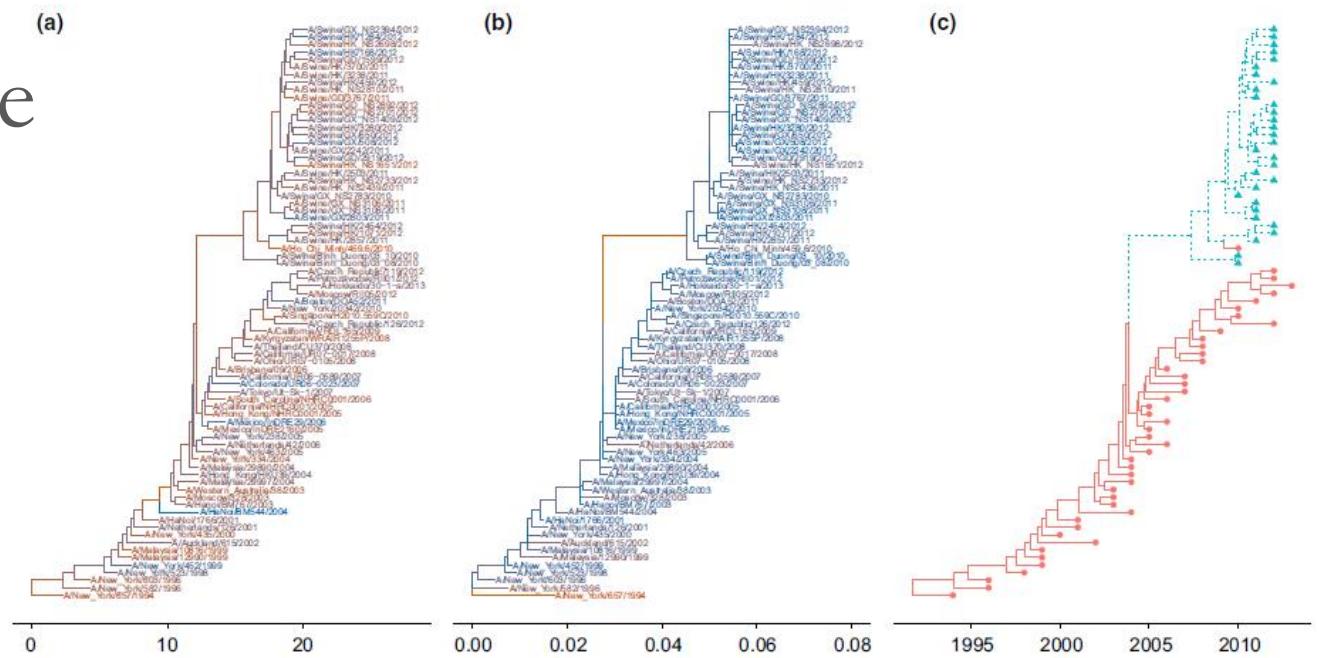


Figure 3.11: **Rotating tree.** A circular/fan layout tree can be rotated by any specific angle.

# Tree manipulation

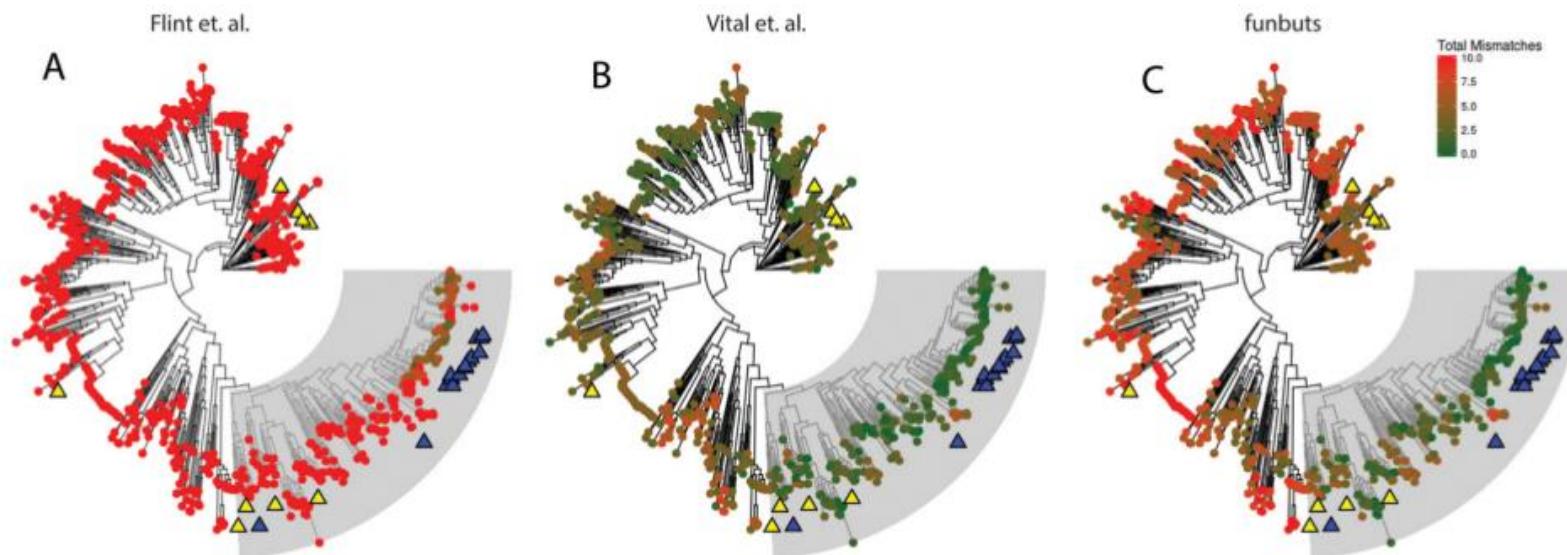
- viewclade
- groupOTU & groupClade
- scaleClade
- collapse & expand
- rotate & flip
- open\_tree & rotate\_tree
- **rescale\_tree**



**Fig. 3.** Phylogenetic tree of H3 influenza viruses. The tree with branches scaled in time (years from the root) and coloured by substitution rates (a). The tree was rescaled using  $d_N$  as branch lengths and coloured by  $d_N$  values (b). The tree branches were rescaled in time (Gregorian calendar) and were assigned to different groups based on host species of the taxa, by which the branches were annotated in different line types, colours and symbols (c).

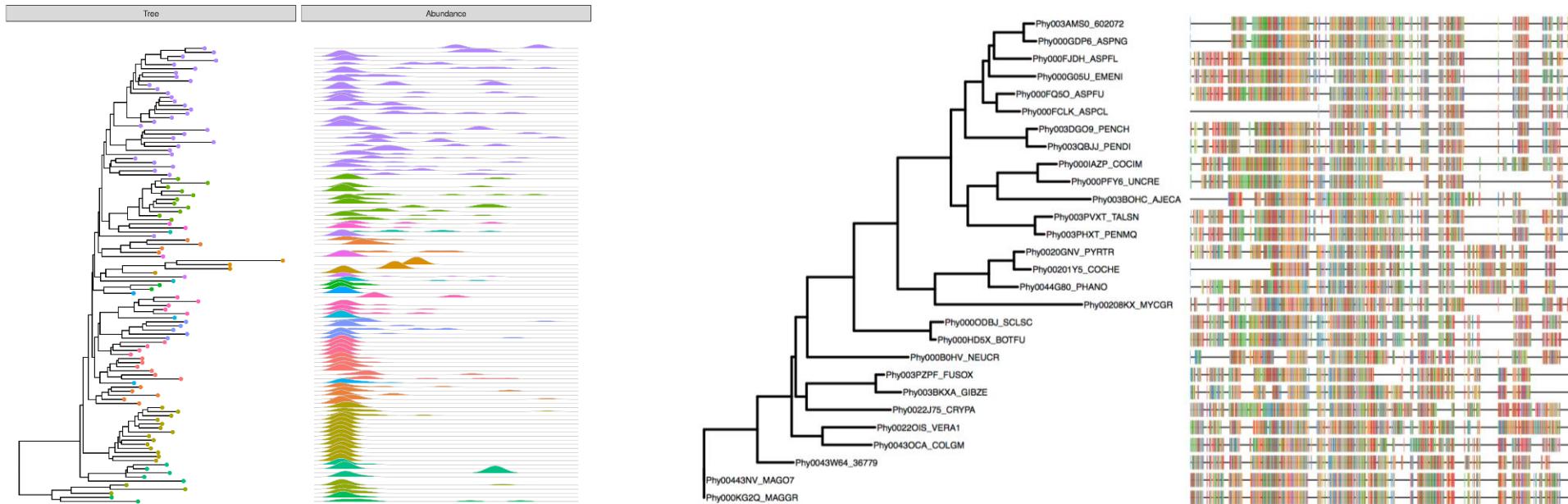
# A tale of two tree-data links

- **Mapping data to the tree structure**
  - Display data on the tree
  - Use data as visual characteristics of tree branches/nodes
- Aligning graph to the tree based on tree structure



# A tale of two tree-data links

- Mapping data to the tree structure
- **Aligning graph to the tree based on tree structure**
  - Dotplot/barplot/violinplot
  - The plot need to be re-ordered based on tree structure



# Mapping data to the tree structure

- The `%<+%` operator

```
library(ggimage)
library(ggtree)
url <- paste0("https://raw.githubusercontent.com/TreeViz/",
              "metastyle/master/design/viz_targets_exercise/")

x <- read.tree(paste0(url, "tree_boots.nwk"))
info <- read.csv(paste0(url, "tip_data.csv"))

p <- ggtree(x) %<+% info + xlim(-.1, 6)
p2 <- p + geom_tiplab(aes(image = imageURL), geom = "image",
                      offset = 2, align = T, size = .16, hjust = 0) +
    geom_tiplab(geom = "label", offset = 1, hjust = .5) +
    geom_tipoint(aes(shape = trophic_habit, color = trophic_habit, size = mass_in_kg)) +
    theme(legend.position = "right") + scale_size_continuous(range = c(3, 10))

d2 <- read.csv(paste0(url, "inode_data.csv"))
p3 <- p2 %<+% d2 + geom_label(aes(label = vernacularName.y, fill = posterior)) +
    scale_fill_gradientn(colors = RColorBrewer::brewer.pal(3, "YlGnBu"))
print(p3)
```

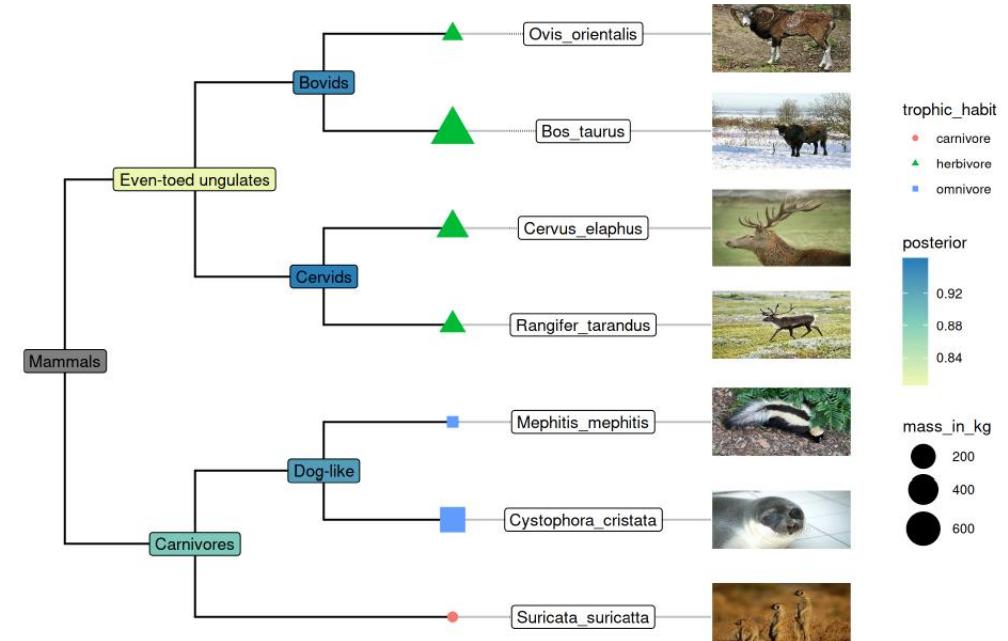


Fig. S1: Example of attaching multiple datasets and labelling taxa with image files

# Aligning graph to the tree based on tree structure

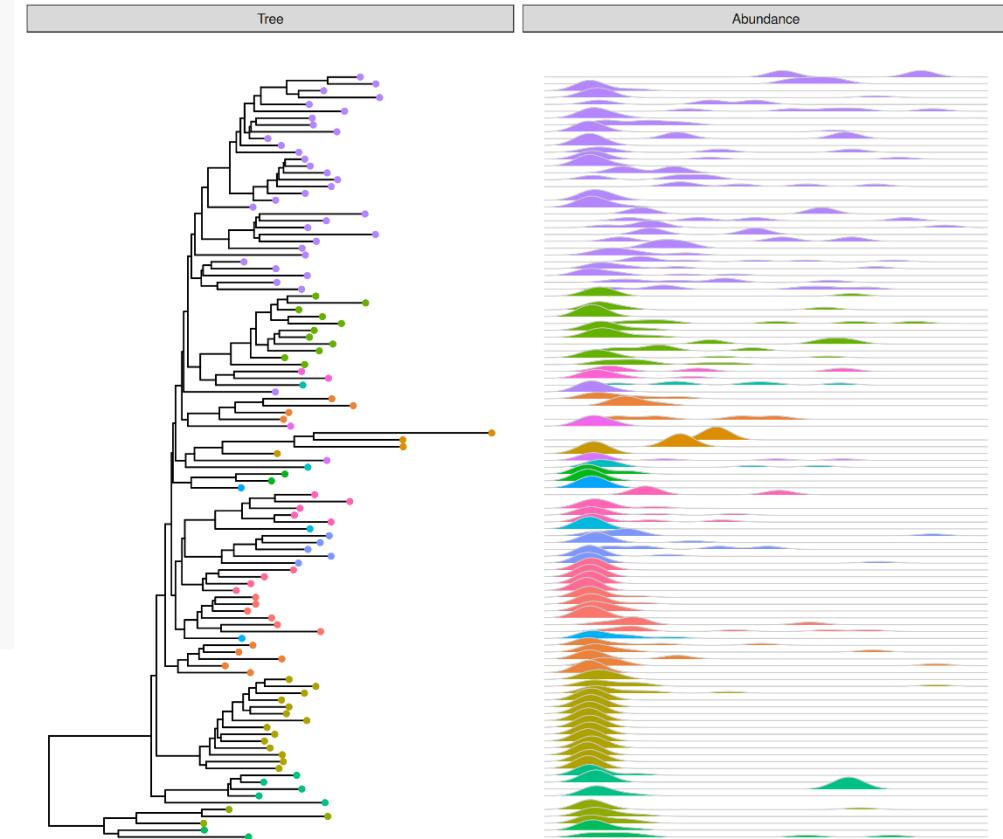
- The *facet\_plot* function

```
data("GlobalPatterns")
GP <- GlobalPatterns
GP <- prune_taxa(taxa_sums(GP) > 1000, GP)
sample_data(GP)$human <- get_variable(GP, "SampleType") %in% c("Feces", "Skin")
mergedGP <- merge_samples(GP, "SampleType")
mergedGP <- rarefy_even_depth(mergedGP, rngseed=394582)
mergedGP <- tax_glom(mergedGP, "Order")

melt_simple <- psmelt(mergedGP) %>%
  filter(Abundance < 120) %>%
  select(OTU, val = Abundance)

p <- ggtree(mergedGP) +
  geom_tippoint(aes(color = Phylum), size = 1.5)

facet_plot(p, panel="Abundance", data=melt_simple,
           geom_density_ridges, mapping = aes(x = val, group = label,
                                             fill = Phylum),
           color = 'grey80', lwd = .3)
```



# Aligning graph to the tree based on tree structure

- The *facet\_plot* function

```
data("GlobalPat")
GP <- GlobalPat
GP <- prune_tax
sample_data(GP)
mergedGP <- mer
mergedGP <- ran
mergedGP <- tax

melt_simple <-
  filter(Abunda
select(OTU, v

p <- ggtr
geom_tippoin

facet_plot(p, p
geom_tippoin

colo
```

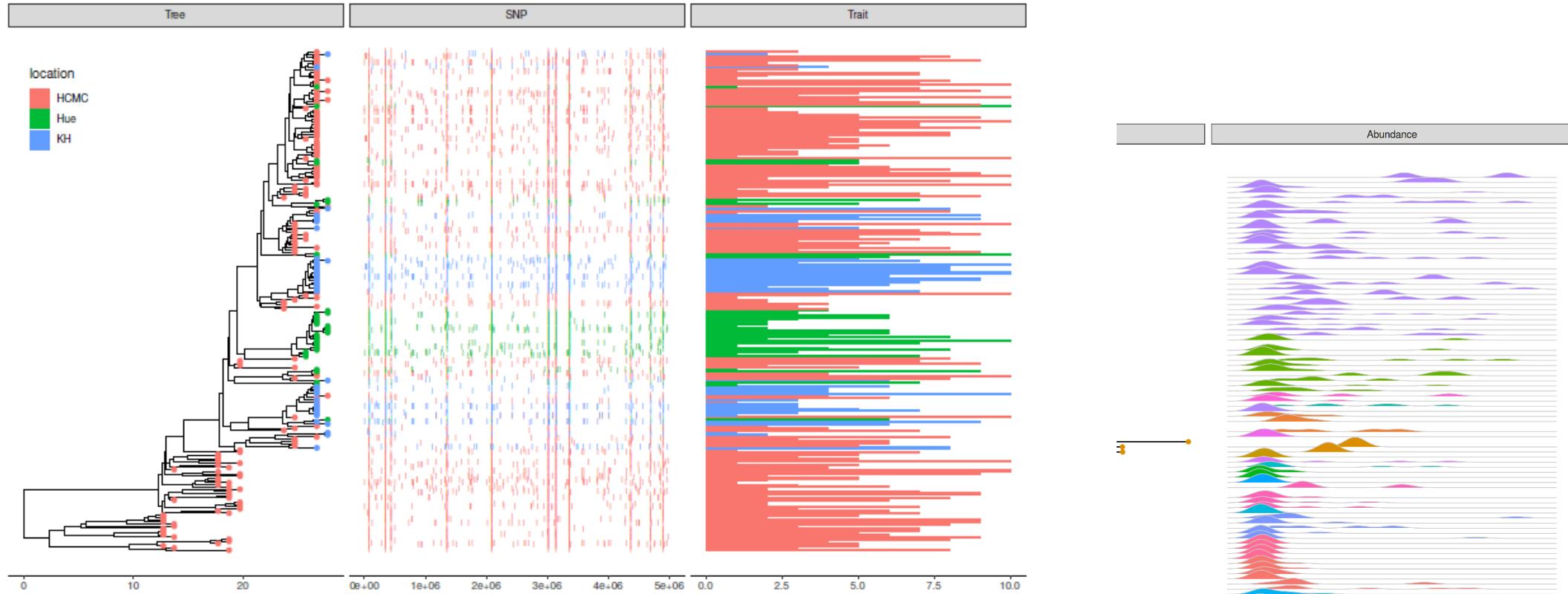
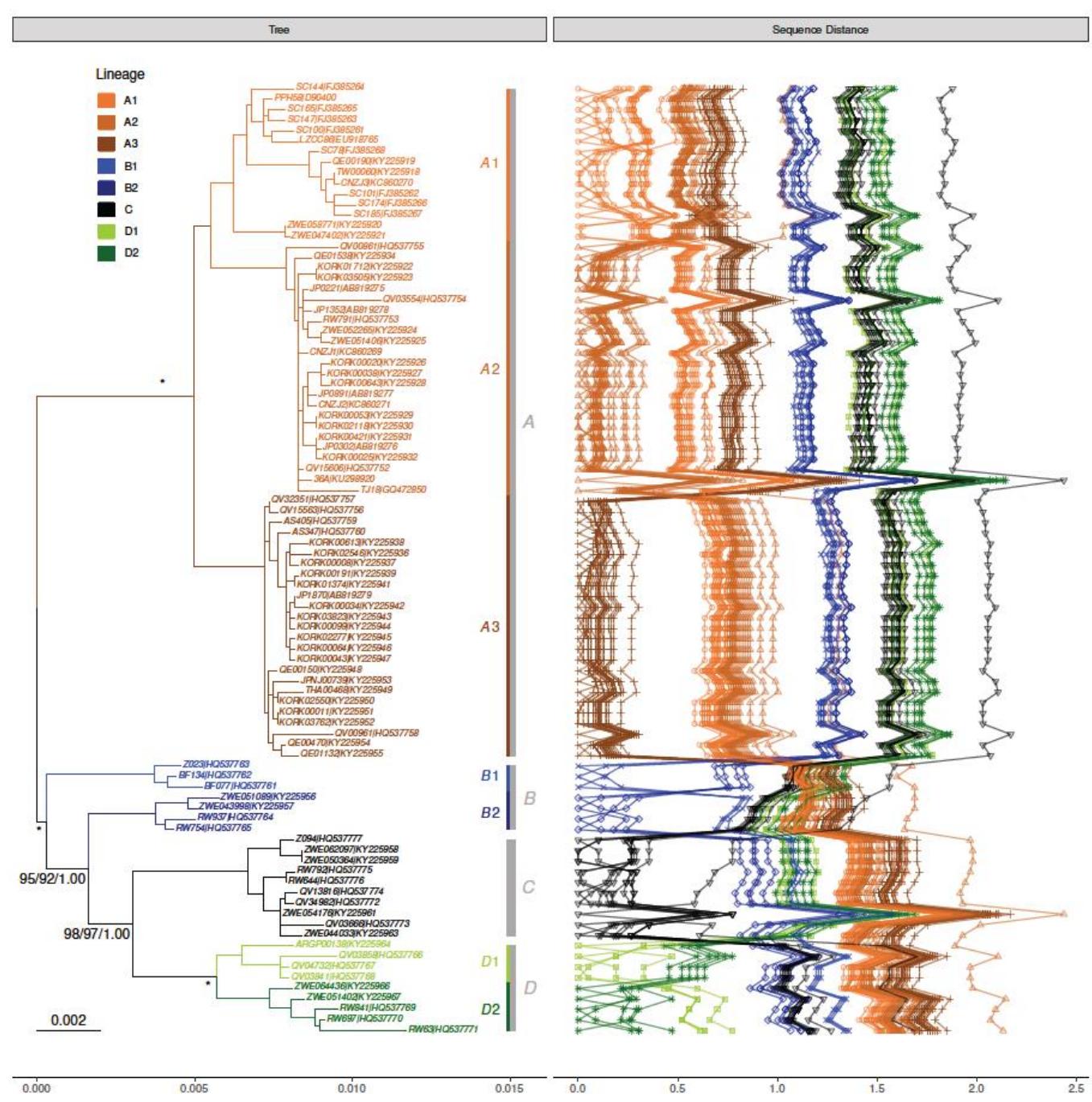
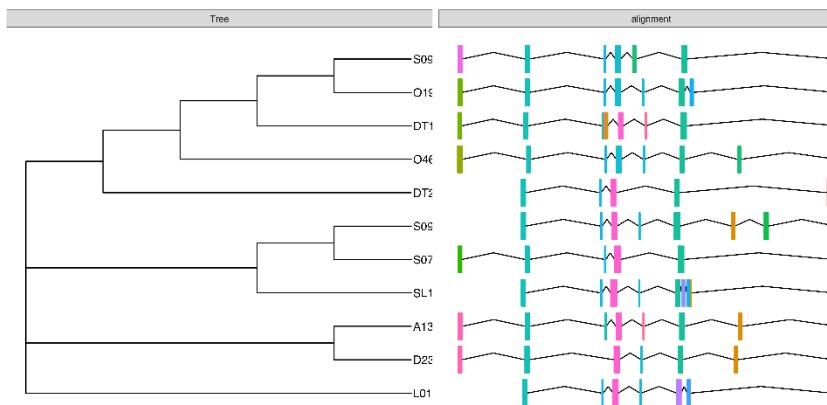
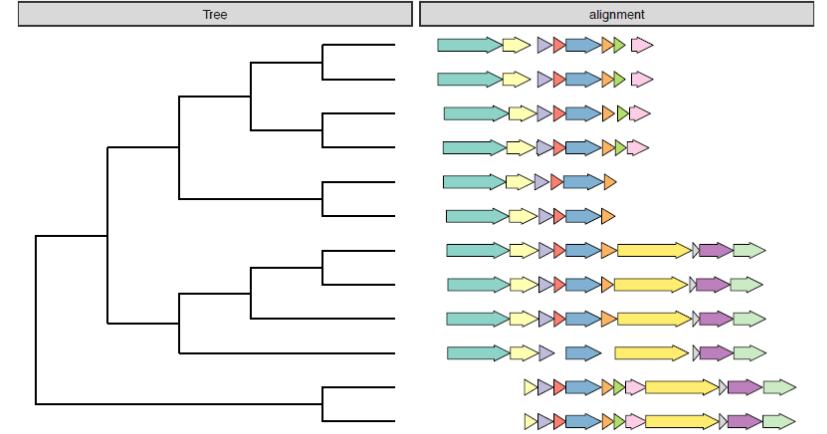
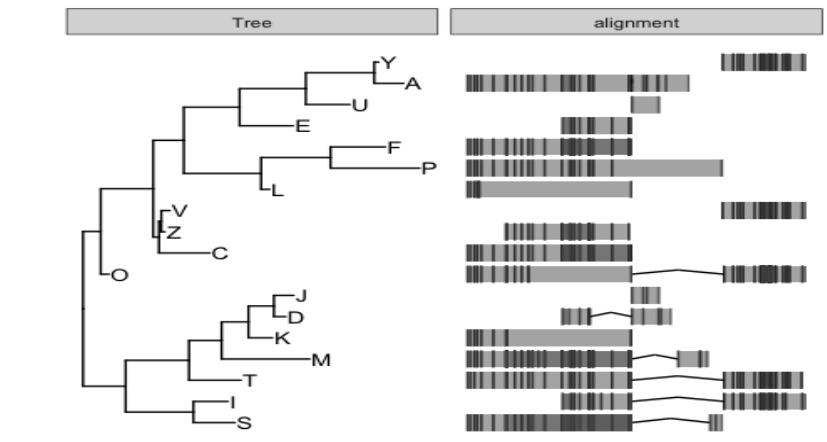


Fig. S6: Example of plotting SNP and trait data using ggtree



Table S1: Geometric layers that supported by 'facet\_plot()'

Package	Geom Layer	Description
ggalt	geom_dumbbell	creates dumbbell charts
ggbio	geom_alignment	shows interval data as alignment
ggfitter	geom_fit_text	shrinks, grows or wraps text to fit inside a defined rectangular area
gggenes	geom_gene_arrow	draws genes as arrows
ggimage	geom_image	visualizes image files
	geom_phylopic	queries image files from phylopic database and visualizes them
	geom_hline	adds horizontal lines
	geom_jitter	adds a small amount of random variation to the location of each point
	geom_label	draws a rectangle behind the text
	geom_point	creates scatterplots
	geom_raster	a high performance special case for all the tiles are the same size
ggplot2	geom_rect	draws rectangle by using the locations of the four corners
	geom_segment	draws a straight line between points
	geom_speke	a polar parameterisation of 'geom_segment()'
	geom_text	adds text to the plot
	geom_tile	draws rectangle by using the center of the tile and its size
	geom_vline	adds vertical lines
	geom_text_repel	adds text to the plot. The text labels repel away from each other and away from the data points
ggrepel	geom_label_repel	draws a rectangle underneath the text. The text labels repel away from each other and away from the data points
	geom_density_ridges	arranges multiple density plots in a staggered fashion
	geom_density_ridges_gradient	works just like 'geom_density_ridges' except that the 'fill' aesthetic can vary along the x axis
	geom_ridgeline	plots the sum of the 'y' and 'height' aesthetics versus 'x', filling the area between 'y' and 'y + height' with a color
	geom_ridgeline_gradient	works just like 'geom_ridgeline' except that the 'fill' aesthetic can vary along the x axis
ggstance	geom_barh	horizontal version of 'geom_bar()'
	geom_boxplot	horizontal version of 'geom_boxplot()'
	geom_crossbarh	horizontal version of 'geom_crossbar()'
	geom_errorbarh	horizontal version of 'geom_errorbarh()'
	geom_histogramh	horizontal version of 'geom_histogram()'
	geom_linerangeh	horizontal version of 'geom_linerange()'
	geom_pointrangeh	horizontal version of 'geom_pointrange()'
	geom_violinh	horizontal version of 'geom_violin()'
ggtree	geom_motif	draws aligned motifs



# Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree

Guangchuang Yu , Tommy Tsan-Yuk Lam, Huachen Zhu, Yi Guan 

*Molecular Biology and Evolution*, Volume 35, Issue 12, 1 December 2018, Pages 3041–3043,

<https://doi.org/10.1093/molbev/msy194>

**Published:** 23 October 2018

“ Cite

Permissions

Share ▾

## Abstract

*Ggtree* is a comprehensive R package for visualizing and annotating phylogenetic trees with associated data. It can also map and visualize associated external data on phylogenies with two general methods. Method 1 allows external data to be mapped on the tree structure and used as visual characteristic in tree and data visualization. Method 2 plots the data with the tree side by side using different geometric functions after reordering the data based on the tree structure. These two methods integrate data with phylogeny for further exploration and comparison in the evolutionary biology context. *Ggtree* is available from <http://www.bioconductor.org/packages/ggtree>.

# Two methods for mapping and visualizing associated data on phylogeny using *ggtree*

- Integrating node/edge data to the tree can be mapped to visual characteristics of the tree or other datasets
- No predefined of input data types or how the data should be plotted in *facet\_plot*
- Combining different *geom* functions to visualize associated data is supported
- Visualizing different datasets on the same panel is supported
- Data integrated by `%<+%` can be used in *facet\_plot*
- Modular design by separating tree visualization, data integration and graph alignment

# The *ggtree* object maintains integrity of tree, annotation data and visual style directives

```
n <- Ntip(nhx_tree)
d <- data.frame(label=as.phylo(nhx_tree)$tip.label,
                 traitA = sample(letters[1:3], n, replace=T),
                 traitB = sample(LETTERS[5:8], n, replace=T))
p <- ggtree(nhx_tree) %<+% d +
      geom_tiplab(aes(color=traitA)) +
      geom_tippoint(aes(shape=traitB), color='firebrick') +
      theme(legend.position="right")
print(p)
```

```
## extract tree from graphic object
tree <- as.treedata(p)

## evolutionary trait also included
get.fields(tree)
```

```
[1] "S"      "D"      "B"      "traitA" "traitB"
```

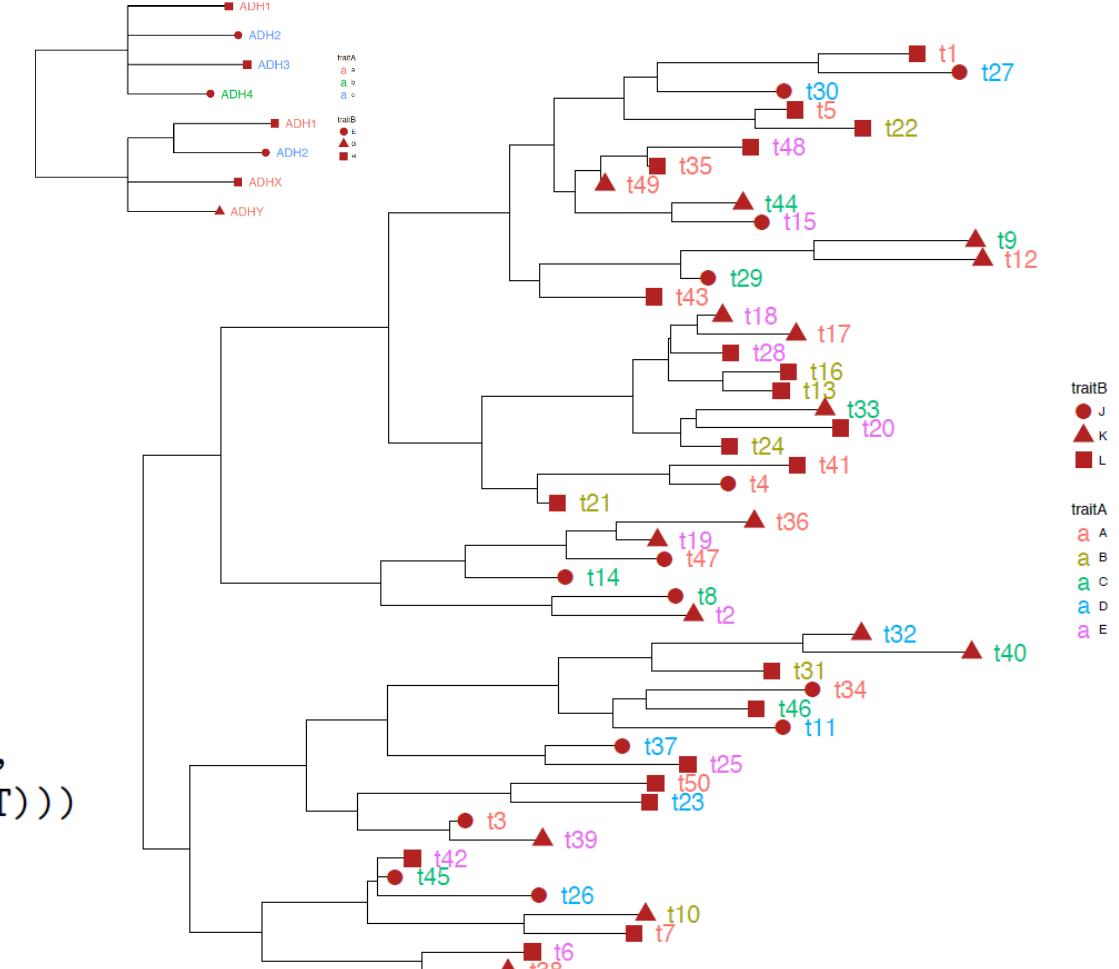
```
## convert graphic object to NEWICK text
write.tree(as.phylo(p))
```

```
((ADHY:0.1,ADHX:0.12,(ADH2:0.1,ADH1:0.11):0.05):0.1,
 (ADH4:0.09,ADH3:0.13,ADH2:0.12,ADH1:0.11):0.1);
```

# The *ggtree* object maintains integrity of tree, annotation data and visual style directives

```
n <- Ntip(nhx_tree)
d <- data.frame(label=as.phylo(nhx_tree)$tip.label,
                 traitA = sample(letters[1:3], n, replace=T),
                 traitB = sample(LETTERS[5:8], n, replace=T))
p <- ggtree(nhx_tree) %<+% d +
      geom_tiplab(aes(color=traitA)) +
      geom_tippoint(aes(shape=traitB), color='firebrick') +
      theme(legend.position="right")
print(p)

tree2 <- new("treedata",
             phylo = rtree(50),
             data = data.frame(node = 1:50,
                               traitA = sample(LETTERS[1:5], 50, replace = T),
                               traitB = sample(LETTERS[10:12], 50, replace = T)))
p %<% tree2
```



# ggtree extensions

[Ich14forever / microbiomeViz](#)

Code Issues Pull requests Projects Wiki Insights

Visualize microbiome data with black magic ggtree

microbiome ggplot2 ggtree r visualization

37 commits 1 branch 0 releases 2 contributors

Branch: master New pull request Create new file Upload files Find file Clone or download

Ich14forever Update README.md

Latest commit 1e7f88a on 6 Nov

- R New branch, phyloseq imported optionally 4 months ago
- data add sample data 11 months ago
- man fixed clado.anno and phyloseq support 6 months ago
- vignettes fixed typo 4 months ago
- .Rbuildignore first commit 11 months ago
- .gitignore removed test data 11 months ago
- DESCRIPTION New branch, phyloseq imported optionally 4 months ago
- NAMESPACE New branch, phyloseq imported optionally 4 months ago
- README.md Update README.md 2 months ago
- microbiomeViz.Rproj first commit 11 months ago

[KlausVigo / ggnetworkx](#)

Code Issues Pull requests Projects Wiki Insights

phylogenetic networks using ggplot2 and ggtree

34 commits 1 branch 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find file Clone or download

KlausVigo improve vignette

Latest commit b08c93a on 8 Jul

- R add index for hybrid edges 9 months ago
- man move ape from Depends to Imports 9 months ago
- tests small improvements 9 months ago
- vignettes improve vignette 6 months ago
- .Rbuildignore add travis 9 months ago
- .gitignore initial commit a year ago
- .travis.yml try to fix travis problem 9 months ago
- DESCRIPTION move ape from Depends to Imports 9 months ago
- NAMESPACE small improvements 9 months ago
- README.md Add batches 9 months ago
- ggnetworkx.Rproj Add vignette (so far empty) Biocconductor style 9 months ago

[willgearty / deptime](#)

Code Issues Pull requests Projects Wiki Insights

Plotting Tools for Anyone Working in Deep Time

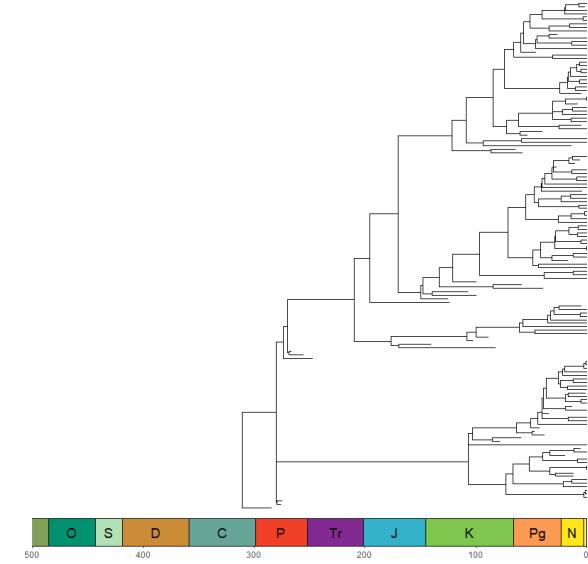
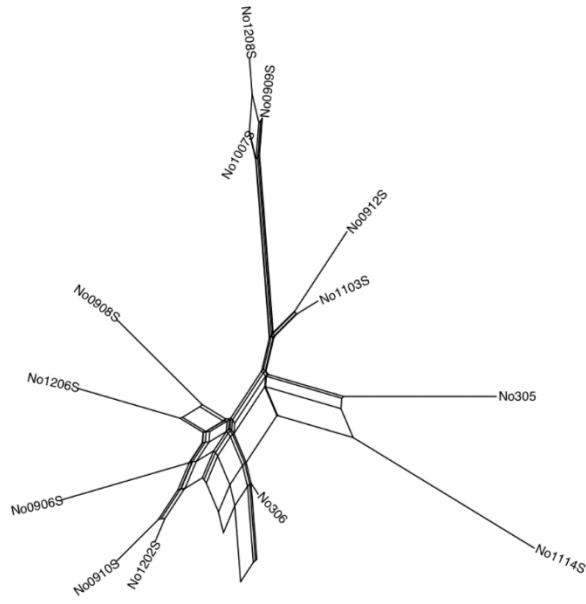
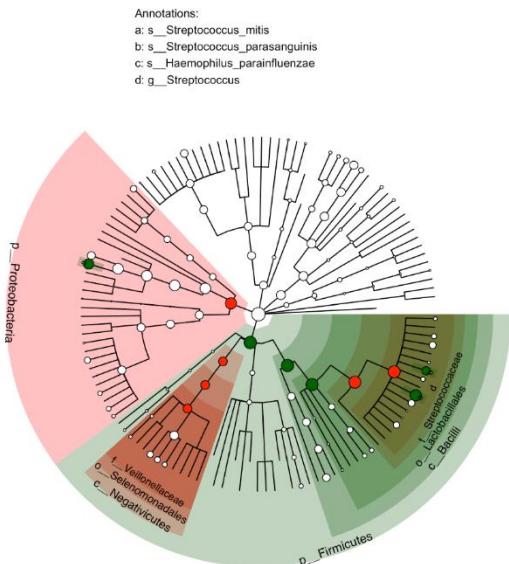
25 commits 1 branch 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find file Clone or download

willgearty Extended period data

Latest commit 37a00cd on 14 Nov

- R Extended period data 2 months ago
- data Extended period data 2 months ago
- images Added example. Rearranged arguments. Added rotation argument. 3 months ago
- man Extended period data 2 months ago
- .Rbuildignore First version of package. Very much still a WIP. 3 months ago
- .gitignore Extended period data 2 months ago
- DESCRIPTION First version of package. Very much still a WIP. 3 months ago
- NAMESPACE First version of package. Very much still a WIP. 3 months ago
- README.md Added more examples 3 months ago
- deptime.Rproj Pulled periods data out of function 3 months ago



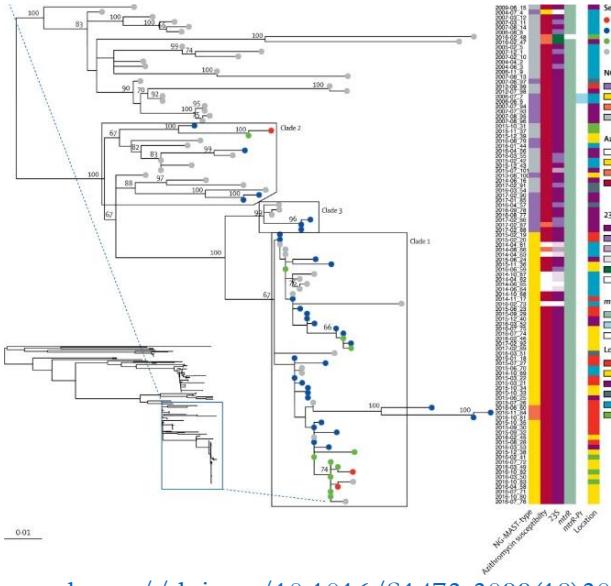
# Summary: It's all about DATA

- **treeio**

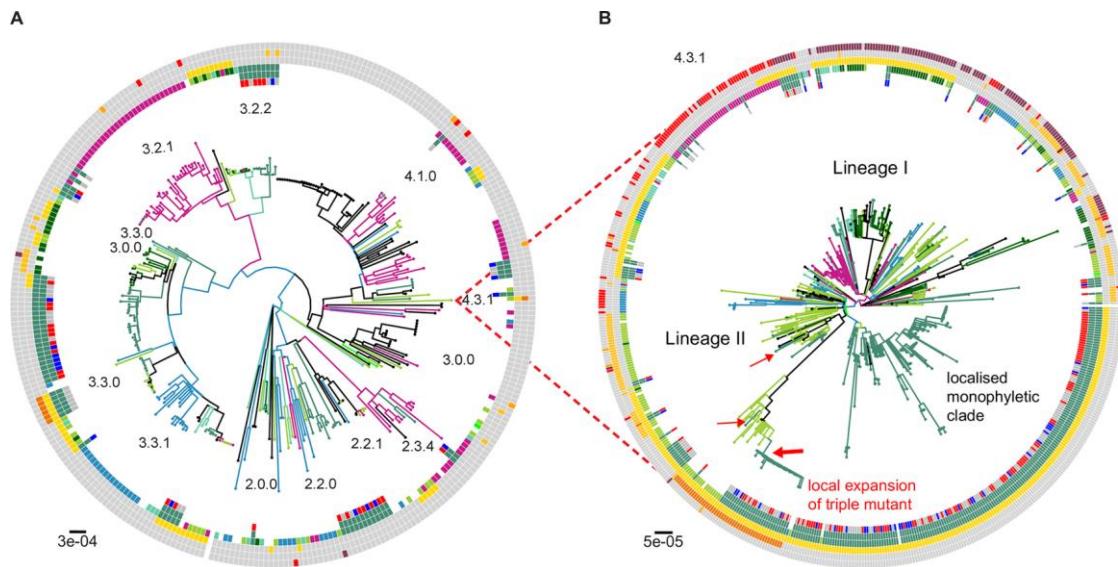
- Importing tree with **data**
- Linking external **data** to tree
- Merging tree **data** from difference sources
- Exporting tree with **data**

- **ggtree**

- Exploring tree **data** visually
- Annotating tree with associated **data**
- Align graph that presents user **data** to the tree
- Integrity of tree, **data** and visual directives

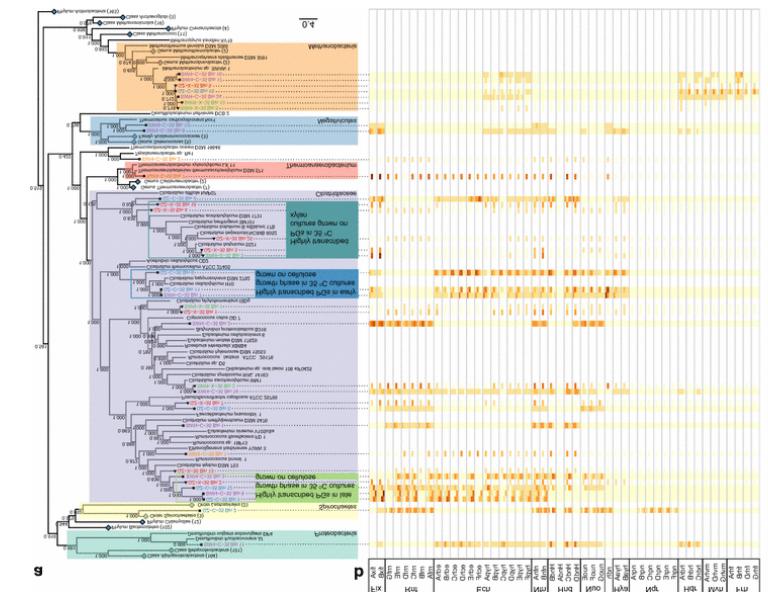


[https://doi.org/10.1016/S1473-3099\(18\)30122-1](https://doi.org/10.1016/S1473-3099(18)30122-1)

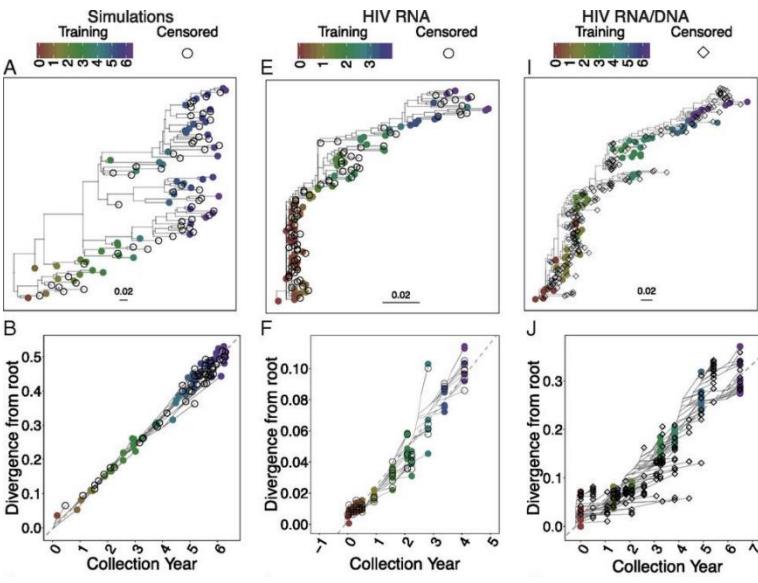


- Rings:**
- 1. Patient status:**
    - Outpatient (this study) (red)
    - Inpatient (this study) (blue)
    - Unknown (previous study) (grey)
  - 2-3. Region:**
    - Nepal (dark green)
    - Bangladesh (medium green)
    - India (light green)
    - Pakistan (yellow-green)
    - Sri Lanka (teal)
    - Africa (blue)
    - Southeast Asia (orange)
    - Other (black)
  - 4-5. QRDR mutations:**
    - No QRDR mutations (grey)
    - gyrA S83F mutation (yellow)
    - 1 additional gyrA/parC mutation (orange)
    - 2 additional gyrA/parC mutations (dark orange)
    - 3 additional gyrA/parC mutations (red)
  - 6. Acquired AMR & MGE:**
    - No AMR/MGE (grey)
    - 1-2 AMR genes (yellow)
    - MDR without IncH1 plasmid (purple)
    - MDR with IncH1 plasmid (red)

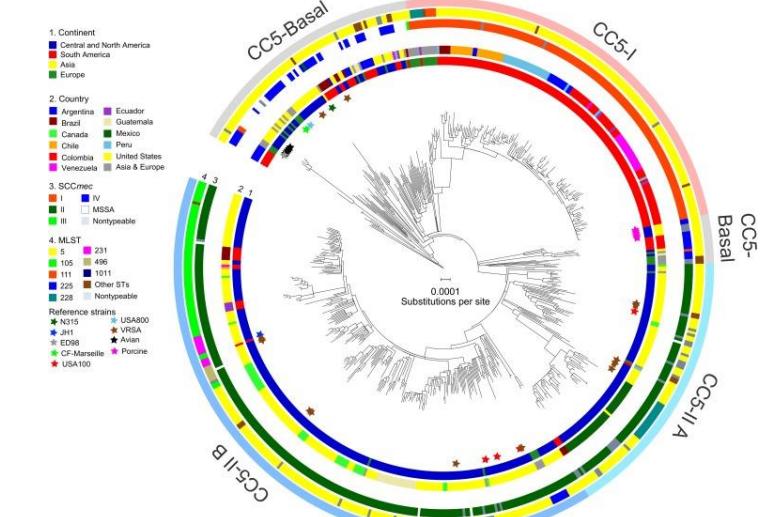
<https://doi.org/10.1371/journal.pntd.0006408>



<https://doi.org/10.1186/s13068-018-1121-0>



<https://doi.org/10.1073/pnas.1802028115>



<https://dx.doi.org/10.3389%2Ffmmicb.2018.01901>

<https://guangchuangyu.github.io/software/ggtree/gallery/>



南方医科大学  
Southern Medical University

Thank You

