

Agenda

- Phylogenetic Tree Data formats
- Tree layouts
- Tree visualization tools
- Visualizing trees with metadata

Phylogenetic Tree Data formats

Phylogenetic tree file formats

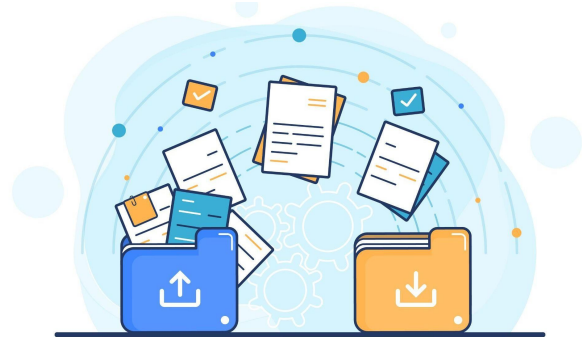
Phylogenetic trees are produced as text files that can have a few different formats:

- Newick (.nwk)- tree topology
- Nexus (.nexus)- tree topology and metadata
- PhyloXML (.xml)- tree topology, metadata and more information about the tree nodes
- JSON (.json)- tree topology, metadata and more information about the tree nodes

Newick format

- Tree topology represented in a parenthesis-based format
- Simple and widely used for its compactness

((A,B),(C,(D,E)));



Nexus format

- Tree topology and associated metadata
- Supports a variety of data types and annotations
- Often used in more complex phylogenetic analyses

#NEXUS

begin taxa;

dimensions

ntax=5; taxa A B

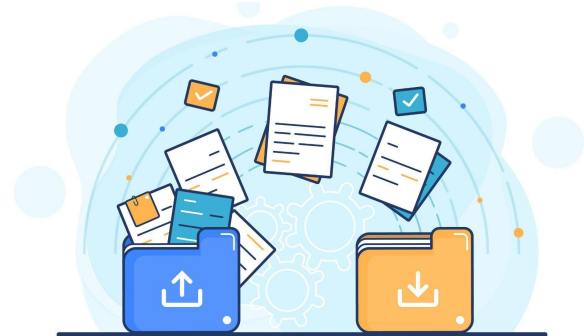
C D E;

end;

begin trees;

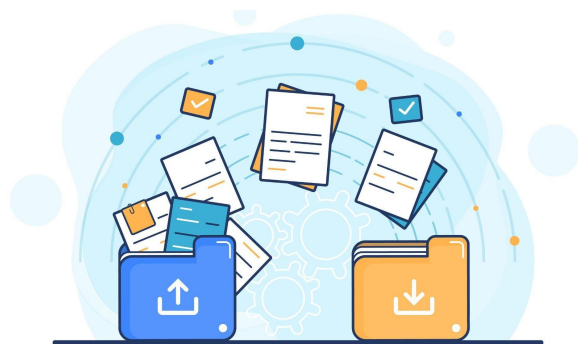
tree example_tree =

((A,B),(C,(D,E))); End;



PhyloXML format

- Tree topology, metadata, and detailed information about tree nodes
- Extensible format allowing for rich annotations and visualization data
- Suitable for complex and detailed phylogenetic data representation



PhyloXML format

```
<phyloxml xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.phyloxml.org  
http://www.phyloxml.org/1.10/phyloxml.xsd" xmlns="http://www.phyloxml.org">
```

```
<phylogeny rooted="true">
```

```
<clade>
```

```
<clade>
```

```
<name>A</name>
```

```
</clade>
```

```
<clade>
```

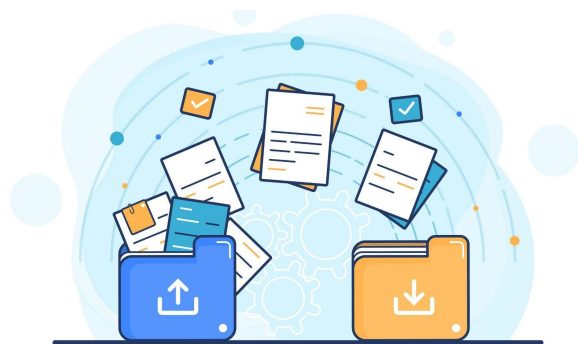
```
.....
```

```
<name>E</name>
```

```
.....
```

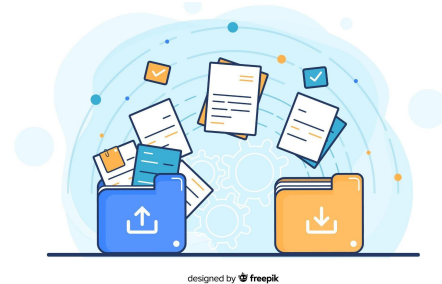
```
</phylogeny>
```

```
</phyloxml>
```



JSON format

- Tree topology and associated metadata in a human-readable format
- Flexible and easy to integrate with web applications and various software tools
- Supports nested structures and complex data relationships
- Often used in modern bioinformatics tools and APIs for interoperability



JSON format

```
{  
  "name": "root", "children": [  
  

```

```
    {  
  

```

```
      "name": "A"  
  

```

```
    },  
  

```

```
    {  
  

```

```
      "name": "B"  
  

```

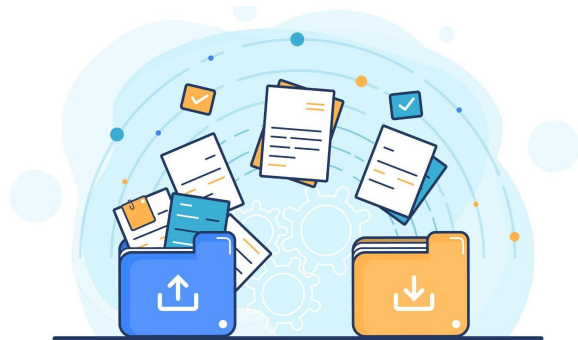
```
    },  
  

```

```
    {  
  

```

```
    ....  
  ]  
}
```



Phylogenetic tree file formats

Note that phylogenetic tree visualization tools are often very specific to their input files

Need to think about your tree visualization tool as you export your tree file

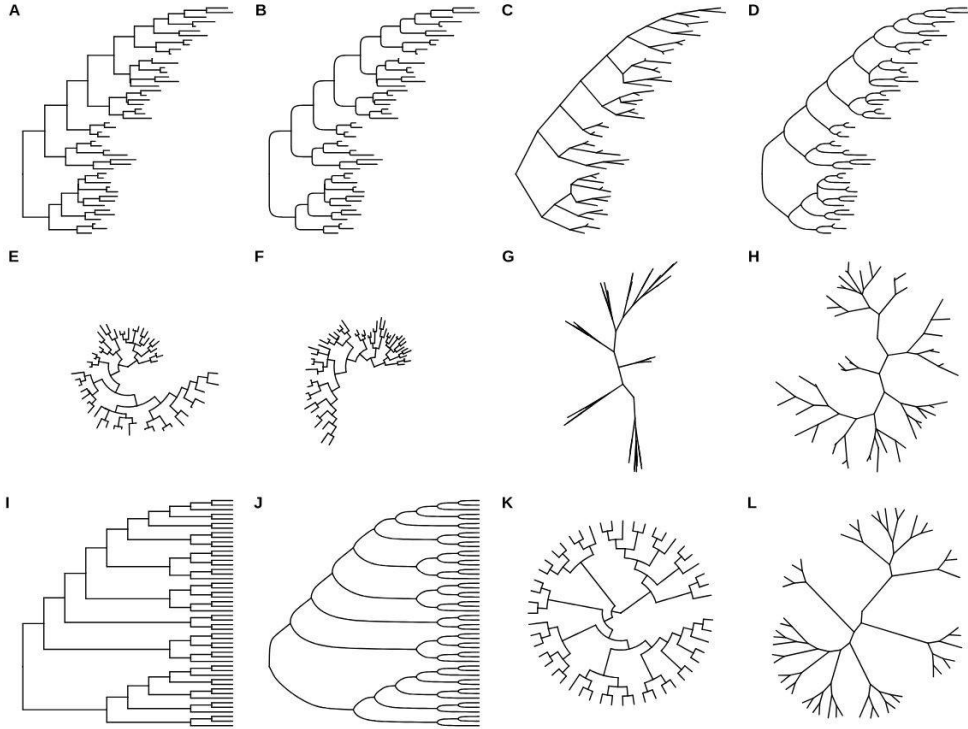
Consider exporting in multiple formats if possible

Phylogenetic tree layouts

Tree Layouts

- There are several types of layouts for tree presentation including **rectangular, rounded rectangular, ellipse, slanted, circular, fan, unrooted (equal angle and daylight), time-scaled, network**, etc.
- Each layout has its specific use cases and advantages, making them suitable for different types of phylogenetic analysis and data presentation.
- Choosing the right format depends on the dataset's complexity, the relationships you want to highlight, and the audience's needs.

Tree Layouts



Phylogram:

A: rectangular, B: rounded rectangular, C: slanted, D: ellipse, E: circular and F: fan

Unrooted:

G: equal-angle and H: daylight method

Cladogram:

I: rectangular, J: ellipse, K: circular, and L: unrooted

Tree Layouts

- **Rectangular**

- Displays the tree with branches at right angles, resembling a hierarchical structure.
- Ideal for presenting evolutionary relationships in a clear, straightforward manner. Commonly used in textbooks and publications.

- **Unrooted**

- Does not show a root; focuses on the relationships between taxa without implying an evolutionary path.
- Employed when the root of the tree is unknown or when relationships need to be shown without implying direction.

Tree Layouts

- **Circular (Radial)**

- Trees are arranged in a circular layout, with the root at the center and branches radiating outward.
- Useful for displaying large trees and highlighting relationships in a compact space.

- **Force-directed (Network)**

- Uses algorithms to position nodes based on their relationships, often resembling a network diagram.
- Ideal for visualizing complex relationships and networks, including recombination events.

- **Time-tree**

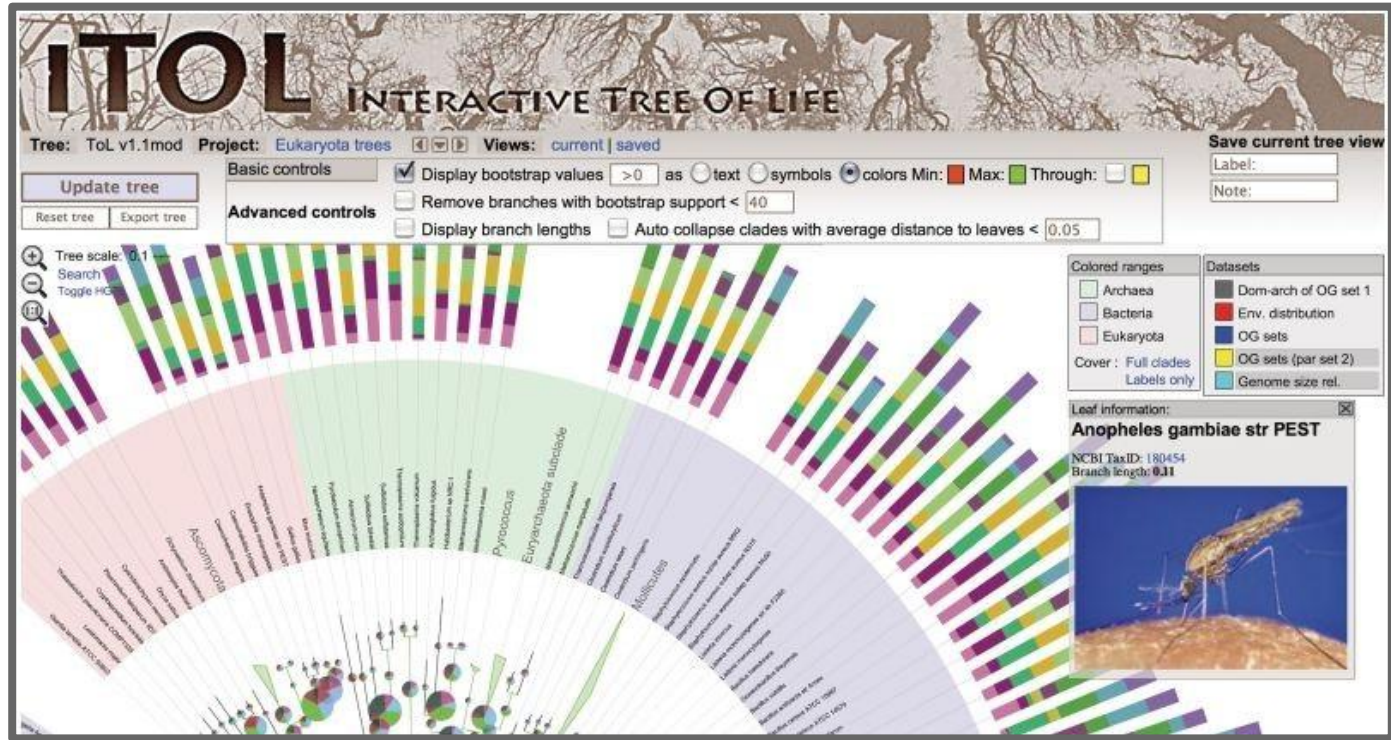
- Branch lengths are proportional to time, showing evolutionary time scales.
- Employed in studies focusing on the timing of evolutionary events.

Phylogenetic tree visualization tools

Visualization platforms...

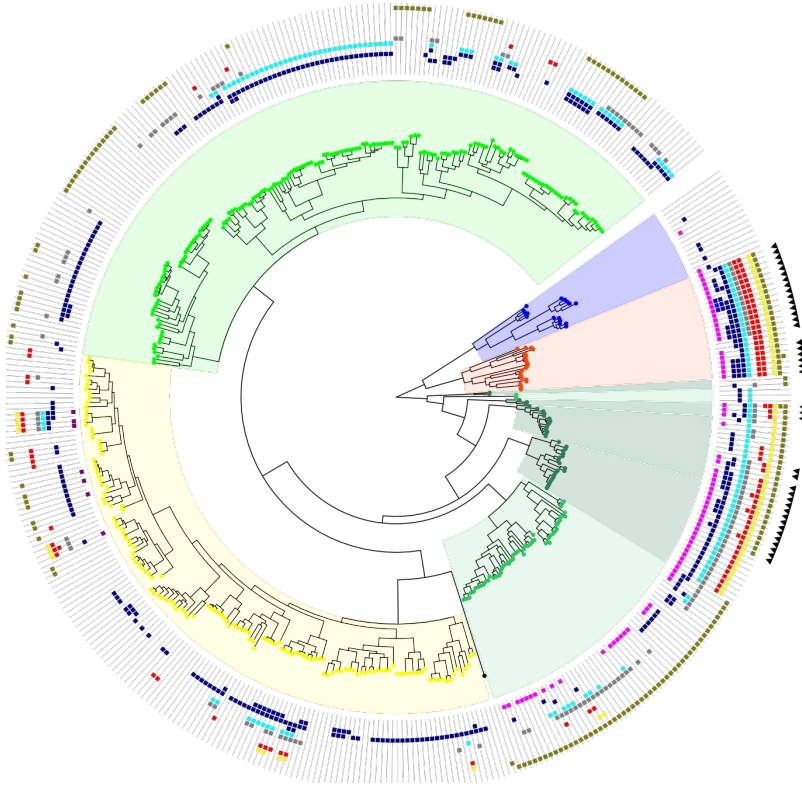
- There are many software packages and web tools that are designed for displaying phylogenetic trees
- Different features and suitability
 - What do you want to communicate?
 - Interactivity
 - Output options
- Diverse forms: online software, desktop software, or libraries
 - Ease of use
 - Compute resources
 - Size of tree (scalability)

iTOL



<https://itol.embl.de/>

iTOL



- Colored dots on the branches are defined using a **branch symbols dataset**.
- There are several **colored ranges** specified.
- Outside the tree, there are two **shape plots**, one using **squares**, and the other **triangles**.

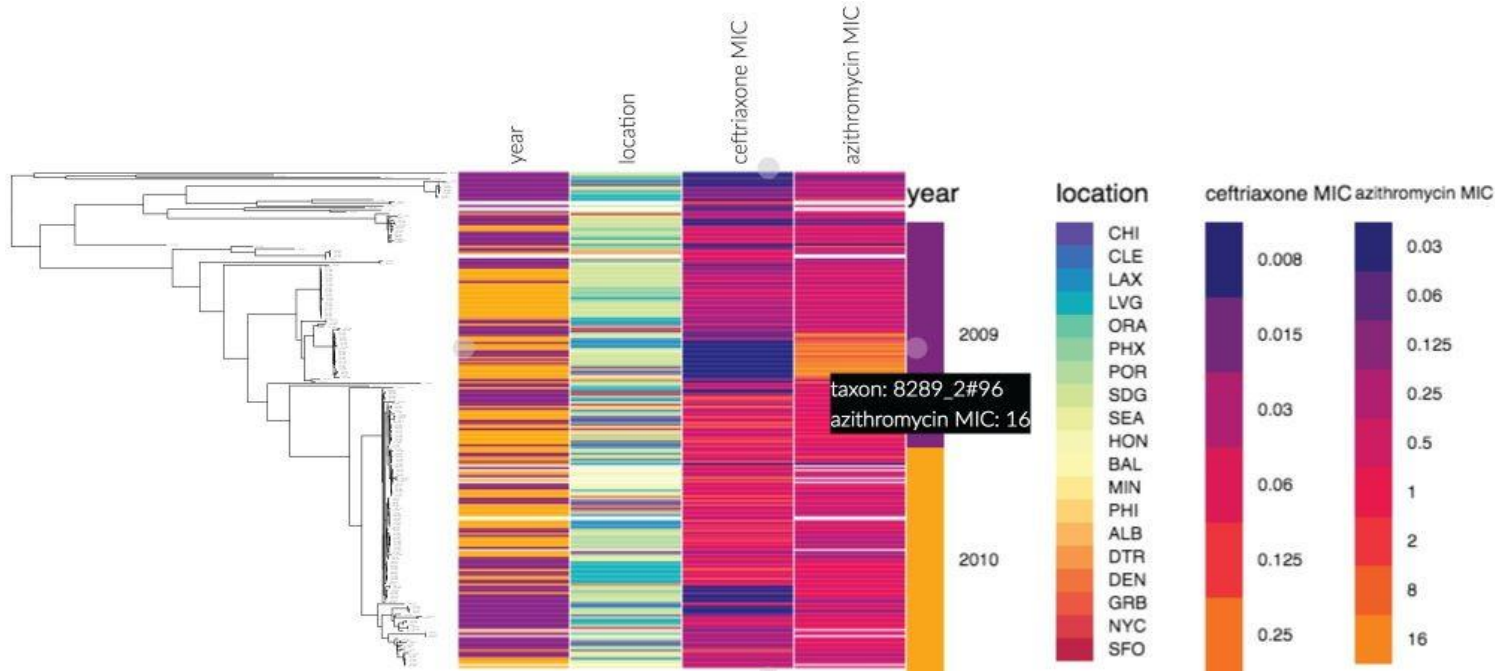
Phandango

Interactive visualization of genome phylogenies



<https://jameshadfield.github.io/phandango/#/>

Phandango: Recommended for visualizing metadata vs tree

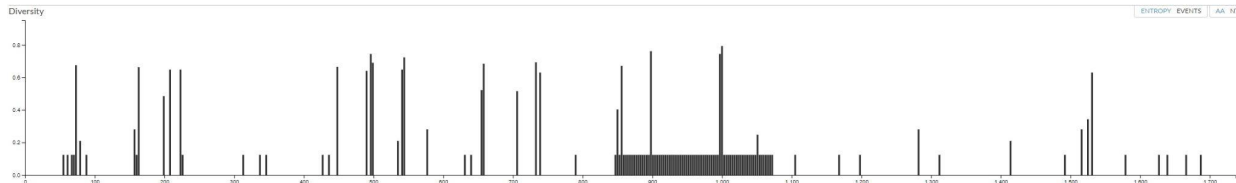
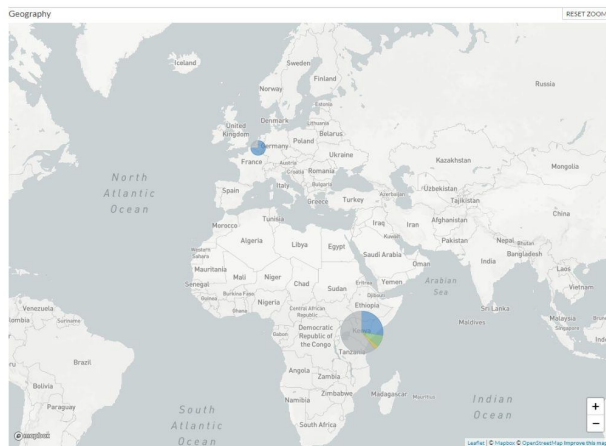
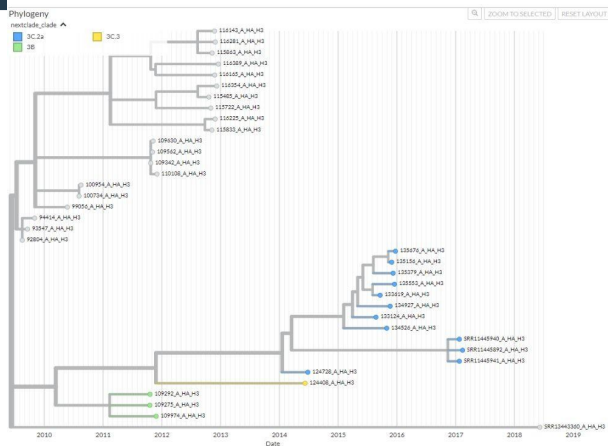


Auspice: Recommended for visualizing tree vs complex data

Real-time tracking of influenza A/H3N2 evolution

Built with [nextstrain/seasonal-flu](#). Maintained by [Jover Lee](#), [Richard Neher](#) and [Trevor Bedford](#). Enabled by data from [GISAID](#)

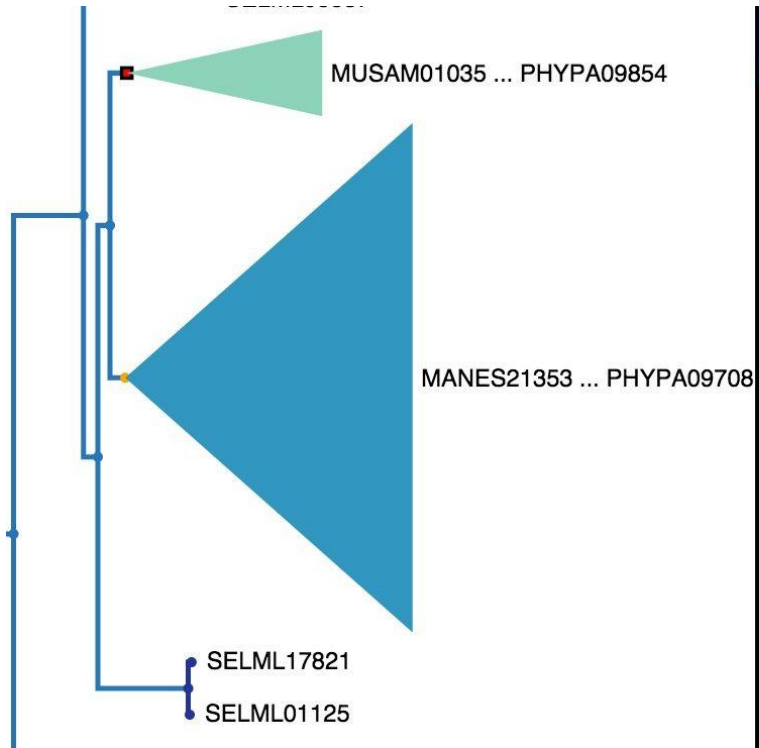
Showing 37 of 37 genomes sampled between Aug 2009 and Jun 2018.



Phylo.io



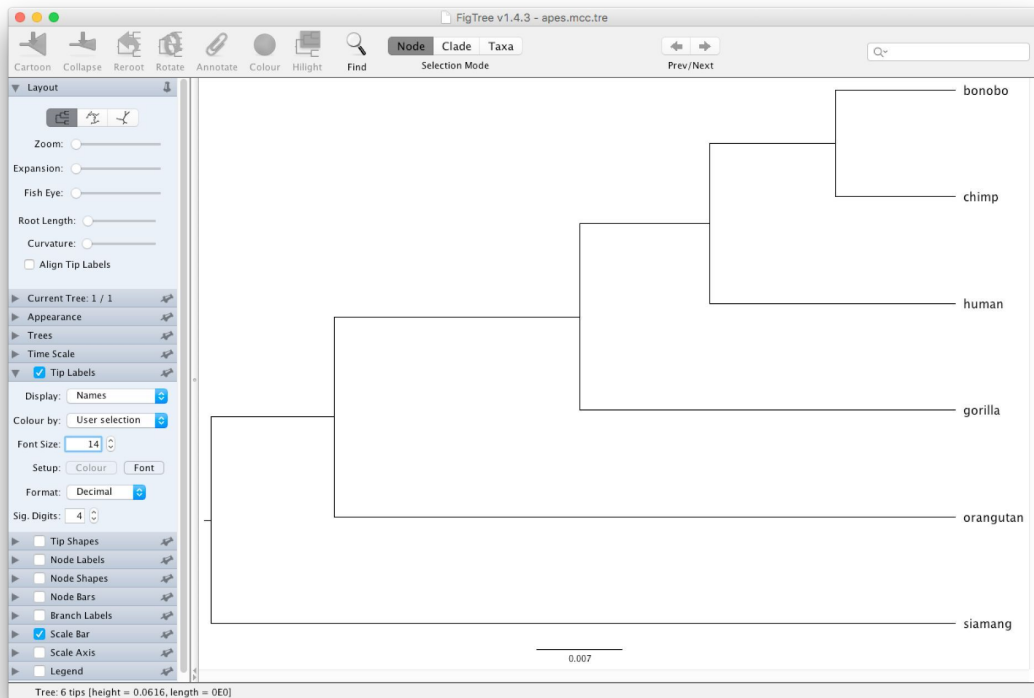
Phylo.io



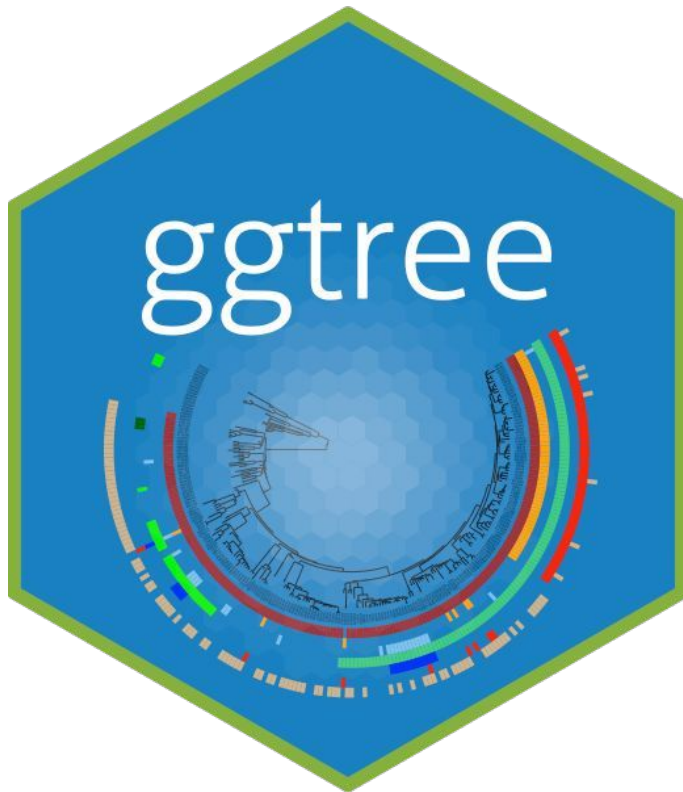
FigTree



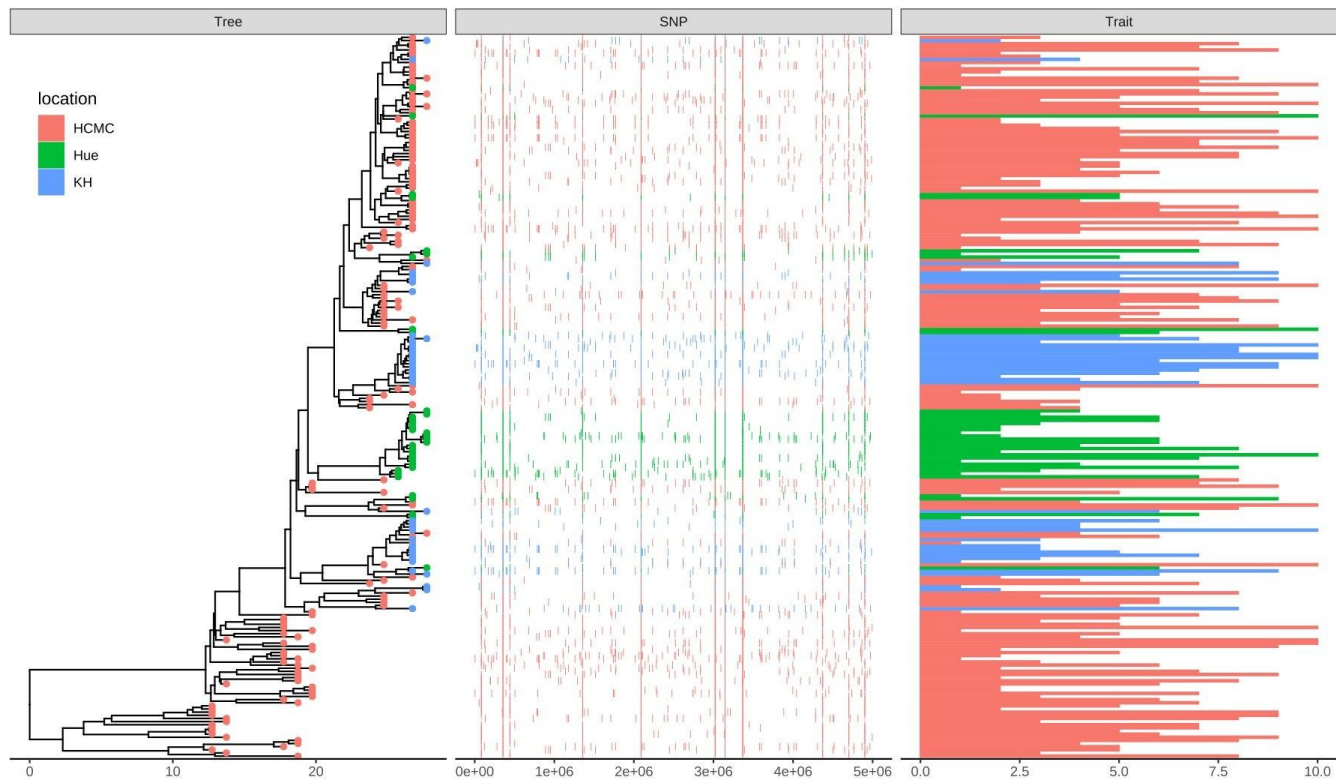
Figtree



ggtree



ggtree



Microreact

Visualization trees with metadata

Why visualize trees with metadata?

- **Enhanced Interpretation:**

- Adding metadata (such as geographic, temporal, or epidemiological data) to phylogenetic trees provides a richer context, aiding in the interpretation of evolutionary relationships and patterns.

- **Communication and Collaboration:**

- More informative and easier to understand for a broad audience, facilitating communication and collaboration among researchers, public health officials, and policymakers.

- **Identification of Patterns:**

- Metadata allows for the identification of patterns that may not be apparent from the tree topology alone. For example, geographic data can reveal the spread of pathogens, while temporal data can show the timeline of evolutionary events.

- **Hypothesis Generation:**

- Metadata-enriched trees can help generate hypotheses about factors influencing evolutionary processes, such as the role of specific environmental conditions or host interactions.

Misconceptions about visualizing trees with metadata?

- **Metadata Automatically Improves Analysis:**
 - Adding metadata to a phylogenetic tree does not guarantee better analysis. The quality and relevance of the metadata are crucial.
- **All Metadata is Equally Important:**
 - Not all types of metadata will be relevant. Select metadata that is directly related to the research question or public health concern being addressed.
- **One Visualization Style Fits All:**
 - Different types of metadata may require different visualization styles. For example, geographic data might be best represented on a map-based tree, while temporal data could benefit from a time-scaled phylogeny.

Misconceptions about visualizing trees with metadata?

- **Visualization Replaces Statistical Analysis:**

- Visualization complements but does not replace the necessity of rigorous statistical analysis for drawing robust conclusions.
- For example, observing a clustering of nodes by geographical location in a phylogenetic tree is not definitive proof of restricted pathogen spread between locations.
- To substantiate such observations, formal phylogenetic-aware tools and statistical methods are required to rigorously analyze and confirm the patterns and underlying processes.

Summary

- Phylogenetic trees are produced as text files that can have a few different formats: Newick, Nexus PhyloXML and JSON
- There are several types of layouts for tree presentation with specific use cases
- There are many software packages and web tools that are designed for displaying phylogenetic trees, with different features and suitability
- Visualizing trees with metadata is great for enhanced interpretation and identification of patterns
- Visualization complements but does not replace rigorous statistical analysis for drawing robust conclusions.