BV-BRC SARS-CoV-2 Genome Reference Tree: Archeaopteryx.js User Guide

Last updated: 05/04/2021

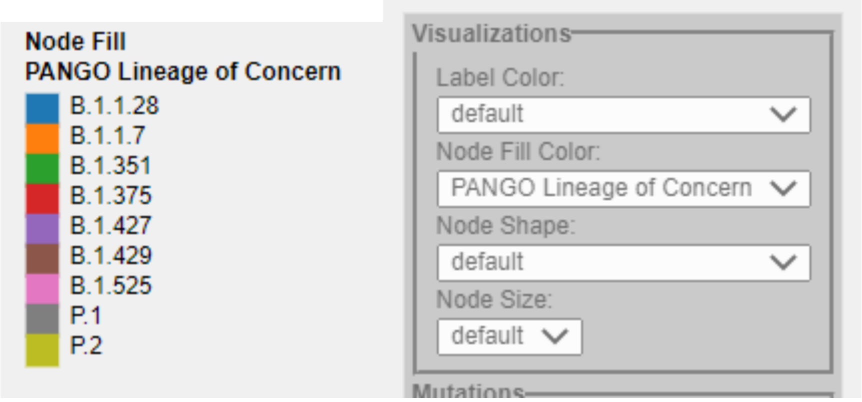
Questions, comments: czmasek@jcvi.org

This document describes how to use Archaeopteryx.js on the Phylogenetic Tree tab on the BV-BRC SARS-CoV-2 Variants and Lineages of Concern resource. More general user documentation is available here.

Initial View

On startup, Archaeopteryx.js uses node fill colors to highlight Lineages of Concern as determined by pangoLEARN (https://cov-lineages.org/pangolin_docs/pangolearn.html), such as B.1.1.7 and P.1 (Figure 1A). Figure 1B shows an enlarged view of the relevant sections.

Figure 1A: Startup view of Archaeopteryx.js on https://bv-brc.org/view/VariantLineage/#view_tab=phlyogeny.



Node Fill
PANGO Lineage of Concern

- 🟦 B.1.1.28
- 🟧 B.1.1.7
- 🟩 B.1.351
- 🟥 B.1.375
- 🟪 B.1.427
- 🟫 B.1.429
- ■ B.1.525
- ■ P.1
- ■ P.2

Visualizations

Label Color: default
Node Fill Color: PANGO Lineage of Concern
Node Shape: default
Node Size: default

Figure 1B: Enlarged view of Node Fill legend and corresponding Visualization selection.

Visualization of Meta-Data

The "Visualizations" menu on the right side (Figure 2) allows users to visualize various metadata fields as colors (node label or node-fill), node-shapes, and node-sizes. Currently, the following metadata fields can be visualized:

- **PANGO Lineage** - Lineages as determined by pangoLEARN, not useful when looking at the entire tree as there are too many lineages to meaningfully visualize (solution: display only a small sub-tree as described below, or only visualize Lineages of Concern)
- **PANGO Lineage Lvl 0** - The first element of the full lineage designation (i.e., "B" for "B.1.1.7")
- **PANGO Lineage Lvl 1** - The first two elements of the full lineage designation (i.e., "B.1" for "B.1.1.7")
- **Host** - Host species from which the virus was isolated (e.g., "Human", "Mink").
- **Country** - Isolation country
- **Year** - Isolation year
- **Region** - Isolation region (e.g., "Africa", "Caribbean")
- **PANGO Lineage of Concern** - Select PANGO lineages which have been deemed concerning.
- **Year_Month** - Isolation year and month

Visualization of Mutations

The "Mutations" menu (Figure 2) allows users to visualize select mutations as either label- or node-color. Mutations are denoted in the form gene:amino acid substitution, e.g., "S:E484K" where "S" stands for Spike protein carrying a substitution of a lysine (K) for a glutamate (E) at position 484.

Visualization of Convergent Mutations

The "Convergent Mutations" menu (Figure 2) allows users to visualize select mutations, which are considered the result of convergent evolution, as either label- or node-color. Mutations are denoted in the form gene:amino acid substitution, e.g., "S:E484K" where "S" stands for Spike protein carrying a substitution of a lysine (K) for a glutamate (E) at position 484

Visualization of Lineages of Concern

Besides visualizing multiple lineages of concern at the same time (as described above), Archaeopteryx.js also allows users to visualize select lineages individually, either as label- or node-color. This is controlled with the "Lineages of Concern" menu (Figure 2).

Visualizations
Label Color: Year_Month
Node Fill Color: default
Node Shape: default
Node Size: default

Mutations
Label Color: default
Node Fill Color: S:E484K

Convergent Mutations
Label Color: default
Node Fill Color: default

Lineages of Concern
Label Color: default
Node Fill Color: default

Figure 2. The Visualizations Menus.

Sub-Trees

To facilitate working with large trees, Archaeopteryx.js allows users to display sub-trees (Figure 3). To display a sub-tree, users click on an internal node (which will be the root-node of the sub-tree) and then click on **Go to subtree** (Figure 4). This can be repeated an arbitrary number of times, to display smaller and smaller subtrees. To return to the full tree, users either click on the root of the subtree and then click on **Return to Supertree** or click on the **R** button in the left-side menu (under "Tools").

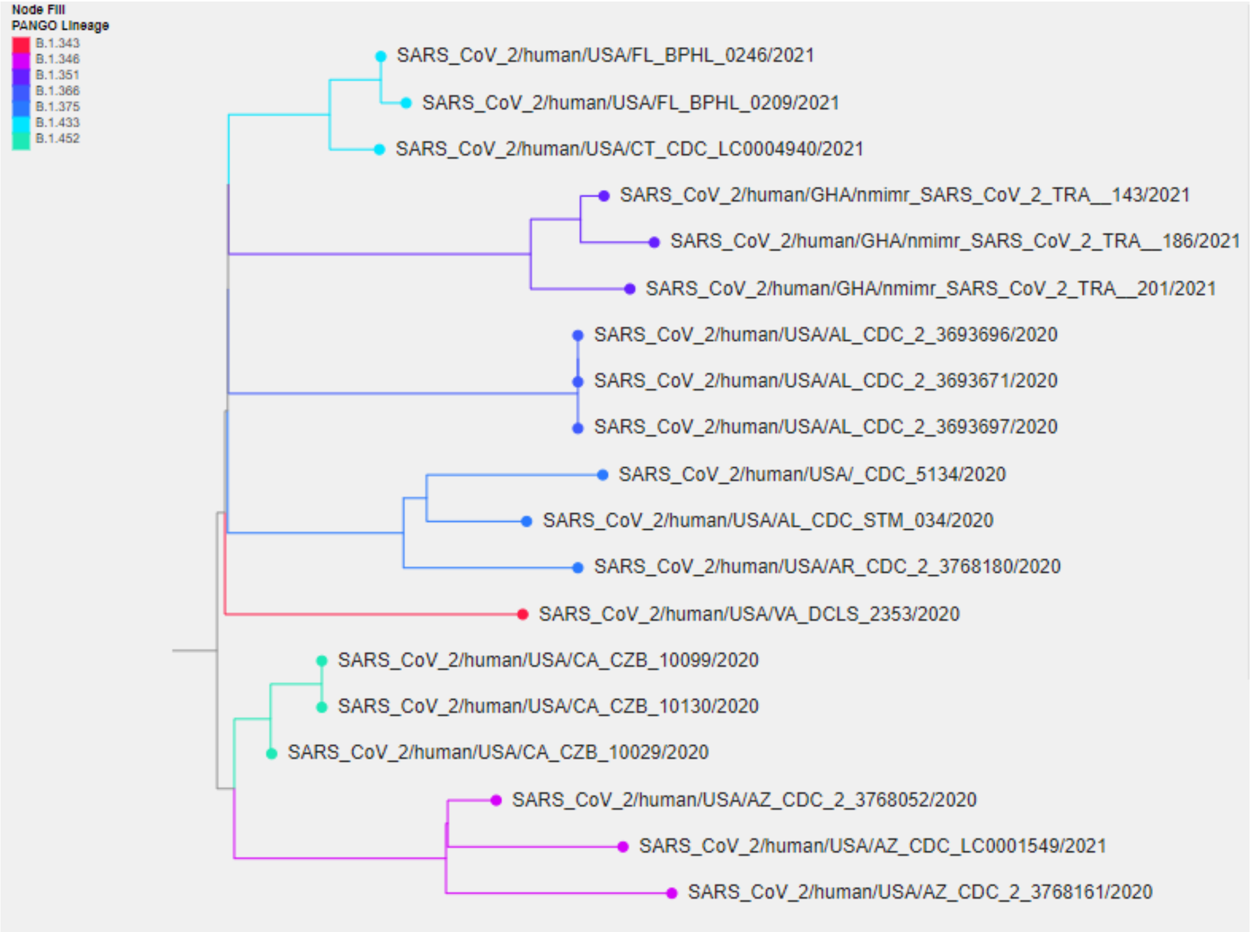


Figure 3. Display of a sub-tree.

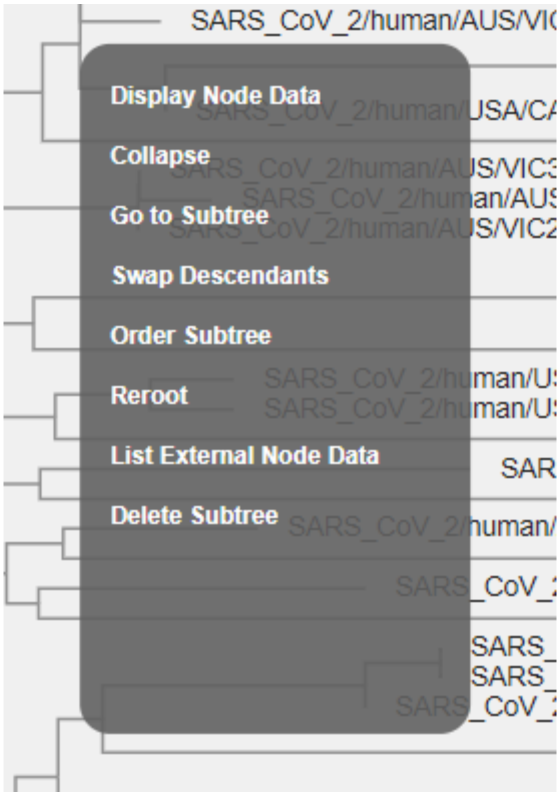


Figure 4. Node menu.

Linking to Sequence Records

Users can inspect the original sequence record associated with external tree nodes by clicking on external nodes, and then clicking on **Access DB**. This will open a new browser tab, containing the sequence record in question.

Search

Users can also search for lineages (or any other data associated with tree nodes) by using Archaeopteryx.js "Search" tools (in the left-side menu).
To search for lineages (such as "B.1.1.7"), the **Wrd** (search for entire words) and **Prp** (search invisible metadata/properties) need to be active (blue) (Figures 5A and 5B).

Search Modes

Search mode options are as follows:
Cas - to search in a case-sensitive manner
Wrd - to match complete words (separated by spaces) only (does not apply to regular expression search)
Neg - to negate (invert) the search results
Reg - to search with regular expressions
Prp - to search hidden properties associated with nodes (such as lineage, country, host)

Logical AND/OR Search

A comma is for logical OR search and a plus sign is used for logical AND search. AND has precedence (is evaluated first) over OR.
Examples

- "2021 + Brazil" only matches nodes with both "2021" and "Brazil" in their name/annotation(s)
- "Brazil, Colombia" matches all nodes with "Brazil" or "Colombia" in their name/annotation(s)
- "2021 + Brazil, 2020 + Colombia" matches all nodes with "2021" and "Brazil" or "2020" and "Colombia" in their name/annotation(s)

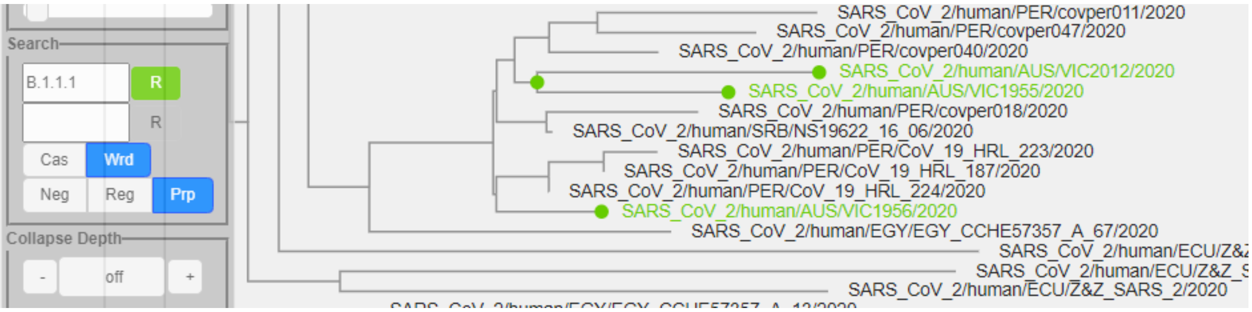


Figure 5A. Searching for lineage B.1.1.1 (found nodes and labels are in green). Note the selected "Wrd" (whole word search) and "Prp" (search properties/metadata) buttons.

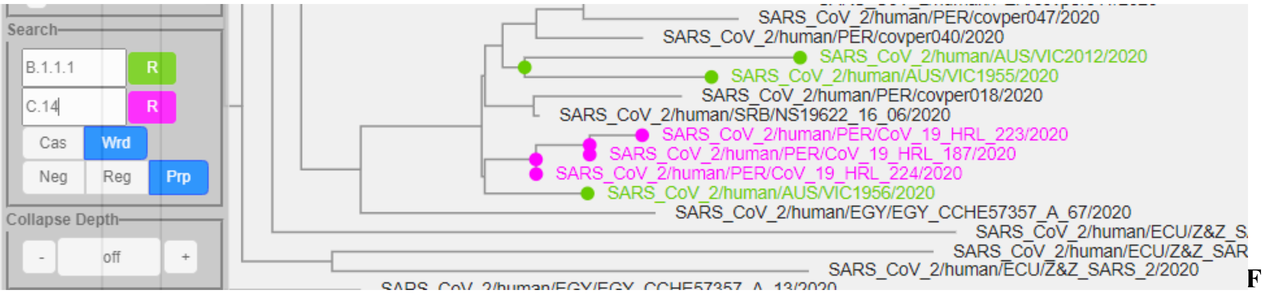


Figure 5B. Searching for lineage B.1.1.1 (found nodes and labels are in green) and lineage C.14 (found nodes are in purple). Note the selected "Wrd" (whole word search) and "Prp" (search properties/metadata) buttons.

Collapsing of Sub-Trees

Another means to facilitate the analysis of large trees is by collapsing of sub-trees. This is accomplished by clicking on an internal node, and then clicking on **Collapse** (Figure 4). Button **U** in the "Tools" menu allows to un-collapse all collapsed sub-trees.

Deletion of Sub-Trees

In contrast to the collapsing of sub-trees, which is reversible, sub-trees can also be deleted which is permanent as far as the displayed tree is concerned, although this does not change the tree stored in the database. It is therefore always possible to return to the original tree by reloading the web-page (Figure 4).

Mouseover

Select information with tree nodes can also be displayed by mouse-over (currently name, lineage, and select mutations).

General Controls

The following describes the standard controls (found in the left control box). Depending on the tree displayed, some buttons might not be present.

Tree Display Types

- **P** - phylogram display (uses branch length values)
- **A** - phylogram display (uses branch length values) with left-aligned labels
- **C** - cladogram display (ignores branch length values)

Display Data

These settings control which data is being shown. In general, only relevant buttons are shown (for example, for a tree which has no internal labels, the "Internal Labels" button is not shown).

- **Node Name** - Show/hide node names
- **Sequence** - Show/hide node sequence information [currently only shows GenBank accession number]
- **Branch Length** - Show/hide branch length values
- **External Labels** - Show/hide external node labels (e.g., node labels, sequence and taxonomic information -- if present)
- **External Nodes** - Show external nodes as shapes (usually circles)
- **Internal Nodes** - Show internal nodes as shapes (usually circles)
- **Node Vis** - Show/hide node visualizations (node colors, shapes, sizes)
- **Branch Vis** - Show/hide node visualizations associated with branches

Zoom

These settings control zoom in horizontal and vertical directions.
General zoom (where everything changes size) can also be achieved with the mouse-wheel.

- **Y+** - Zoom in vertically (Alt+Up or Shift+mousewheel)
- **Y-** - Zoom out vertically (Alt+Down or Shift+mousewheel)
- **X+** - Zoom in horizontally (Alt+Right or Shift+Alt+mousewheel)
- **X-** - Zoom out horizontally (Alt+Left or Shift+Alt+mousewheel)
- **F** - Fit the tree to the display size (Alt+C, Alt+Delete, Home, or Esc to re-position controls as well)

- Alt+plus and Alt+minus to zoom while keeping all font sizes constant
- Shift+Alt+plus and Shift+Alt+minus or Page Up and Page Down or Shift+Ctrl+mousewheel to change all font sizes

Tools

These settings control the tree being displayed:

- **O** - "Order" the entire tree (Alt+O)
- **R** - Return to the super-tree, if in sub-tree (Alt+R)
- **U** – Uncollapse all, if collapse sub-trees present (Alt+U)
- **M** - Midpoint re-root the tree, if tree is re-rootable (uncollapses as well) (Alt+M) [not active for SARS-CoV-2 reference tree]

Size Control

These sliders control the size of various elements:

- **External label size** - Controls the size of the external label fonts
- **Branch label size** - Controls the size of the fonts for confidence and branch lengths
- **Node size** - Controls the size of the external and internal node shapes (if turned on with "External Nodes" and "Internal Nodes")
- **Branch width** - Controls the branch widths (if not set in the tree itself)

Collapse Depth

Collapse sub-tree using a given maximum node depth.

Collapse Feature

Collapse sub-trees based on shared metadata values.

Download

The following formats are available for download:

- PNG image
- SVG vector image
- phyloXML tree file (see http://www.phyloxml.org/)
- Newick tree file

Node Actions

(Left-) clicking on nodes allows to (not all actions are available at all times, see Figure 4):

- Display Node Data
- Collapse
- Uncollapse
- Uncollapse All
- Go to Subtree
- Return to Supertree
- Swap Descendants
- Order Subtree
- Reroot [not active for SARS-CoV-2 reference tree]
- List External Node Data
- Delete Subtree/External Node