# 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 <u>Phylogenetic Tree tab</u> on the <u>BV-BRC</u> <u>SARS-CoV-2 Variants and Lineages of Concern resource</u>. More general user documentation is available <u>here</u>.

# Initial View

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

	Node Fill	SARS CoV 2/human/GHA/nmimr SARS CoV 2 TRA 176/2021	Visualizations
Archaeopteryx.js 1.8.5b6	PANGO Lineage of Concern B.1.1.28	SARS CoV 2/human/USA/UT 08378/2020	Label Color:
P A C	B.1.1.7	SARS CoV_2human/GHA/mminr_SARS CoV_2 TRA_188/2021	default V
Display Data	B.1.351 B.1.375	SARS CoV 2/humanUSA/CA CZE 14789/2020	Node Fill Color:
	B.1.427	SARS_CoV_2human/USA/CA_QDX_2021/2020	PANGO Lineage of Concern V Node Shape:
Node Name	B.1.429 B.1.525	SARS Col/ 2/mumanuSAAL_COC_07312020 SARS Col/ 2/mumanuSAAAL_GOC_07312020	Node Shape:
Sequence	P.1 P.2	SARS CoV 2/human/USA/MI QDX 305/2020	Node Size:
Branch Length	P.2	SARS_CoV_2/human/JAM/JM_GDC_7266/2020 SARS_CoV_2/human/JAM/A_S2847/2020	default 💙
External Labels		SARS CoV 2 human USAWA S2847/2020 SARS_CoV_2 human USA/1500xF7/2020	Mutations
External Nodes		SARS_CoV_2/human/USA/SEARCH_0817_SAN/2020 SARS_CoV_2/human/USA/AK_PHL470/2020	Label Color:
Internal Nodes		SARS CoV 2/human/USA/AZ CDC LC0009211/2021	default 💙
Node Vis		5AR5_00/27/JumanUSAAK152/2020	Node Fill Color:
Branch Vis		SARS_CoV_2/human/MLT/Sliema_2/2020	
Z00m		SARS_COV_2humaniUSANL_UVV_122/0200 SARS_COV_2humaniUSANA_DCLS_0223/2020	Convergent Mutations
		SARS_CoV_2/Homo_saniens/HKG/Case5128/2020	default V
Y4		SAR5 CoV 20umaniNUnGBRC 30202020 SAR5 CoV 20umaniNUnGBRC 30202020	Node Fill Color:
X- F X+		SARS CoV_2humanUSANA_S2802/2020 SARS CoV_2humanUSANCA_C28_2522/2020	default 🗸
Y.		SARS_COV_2human/MUS/MIC138/2020 SARS_COV_2human/SACATOR_2004200	Lineages of Concern-
Tools		SARS CoV 2/human/DEU/NRW 34/2020	Label Color:
O R		SARS CoV 2 human POLPL WG 432020 SARS CoV 2 human POLPL WG 5052020	default V Node Fill Color:
U M		SARS_CoV_2human.WEX/SEARCH_1708_TU/2020 SARS_CoV_2human.USA/CA_QDX_3851/2021	default V
		SARS CoV 2/human/USA/1436xEB/2020	Vis Legend-
External label size:		SARS CoV 2/numan FD-2/DC-840/020	Show Dir
Branch label size:		SARS_CoV_2/human/USA/N0220 SARS_CoV_2/human/USA/N_CDC_2_3693765/2020	
Node size:		ARS CoV 21humanUSACIN, DCC 2, 3087882020	< R >
Branch width:		SARS_CoV_2/human/BGD/GC/	
		- SARS_COV_2human USALC_225_1082/2020 SARS_COV_2human USALC_225_101/2020	
Search-		SARS CoV 2/human/USA/FHCRC 14109/2020	
R		SAPS_CAV_2humanUSA/TX_CDX_38342020	
R		SARS_C0V_2humanUSAFHCRC_14117/2020 SARS_C0V_2human/POL/PL_P23/2020	
Cas Wrd		SARS_COV_2human/AUS/VIC2214/2020 SARS_COV_2human/US/AZ_CDC_LC00000002020	
Neg Reg Prp		SARS_COV_2thuman/USA/CA_C2B_19929/2020 SARS_COV_2thuman/USA/FL_CDC_STM_00000828/2021	
		SARS CoV 2/human/USA/FL CDC LC0012946/2021	
Collapse Depth		SAR2-CV/2/hmmanVSACO_COPHE_210031882021 SAR2-CV2 2/hmmanVSACA C2E 118072020	
- off +		SARS CoV 2/human/USA/SEARCH 0410 SAN/2020	
Collapse Feature		SARS_DOV_2human/AUS/NC_CEA72020 SARS_COV_2human/ND/GERC_376a2020	
off 🗸		SARS CoV 2/human PER/CoV 19 HRL 187/2020	
		SAR5_C0V_2/humanUSA/C0VID19_AK141/2020 SAR5_C0V_2/humanUSA/CA_0DX_1417/2020	
Download		SARS_CoV_2/human/USA/CA_QDX_38/2020	
PNG ¥		SARS_GAV_ZhumanUSAFL_BPHL_21212020 SARS_GAV_ZhumanUSAFL_0DX_1082020	
		SARS_CoV_2human/AUS/TAS133/2020	

Figure 1A: Startup view of Archaeopteryx.js on <a href="https://bv-brc.org/view/VariantLineage/#view-tab=phlyogeny">https://bv-brc.org/view/VariantLineage/#view-tab=phlyogeny</a>.

Node Fill PANGO Lineage of Concern	Visualizations		
B.1.1.28	default V		
B.1.1.7 B.1.351	Node Fill Color:		
B.1.301 B.1.375	PANGO Lineage of Concern 🗸		
B.1.427	Node Shape:		
B.1.429 B.1.525	default 🗸 🗸		
P.1	Node Size:		
P.2	default 🗸		
	Mutations		

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 nodecolor. 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 nodecolor. 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:						
Node Fill Color:						
S:E484K V						
3.L404K V						
Convergent Mutations						
Label Color:						
default 🗸						
Node Fill Color:						
default 🗸						
Lineages of Concern						
Label Color:						
default V						
Node Fill Color:						

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.

	SAR	S_CoV	_2/human//	AUS/VIC
	Display Nod	le Data	_2/human/	JSA/C/
	Collapse			
	Go to Subtr	~~ ~		
	Swap Desce	endants		
	Order Subtr	ee		
_	Reroot			
				man/0
	List Externa			SAR
	List Externa Delete Subt	I Node C	)ata	SAR
		I Node C	D <b>ata</b> (S_CoV_2/	SAR
		I Node C	D <b>ata</b> (S_CoV_2/	SAR human/

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.



**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 <u>not</u> 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 <u>http://www.phyloxml.org/</u>)
- 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