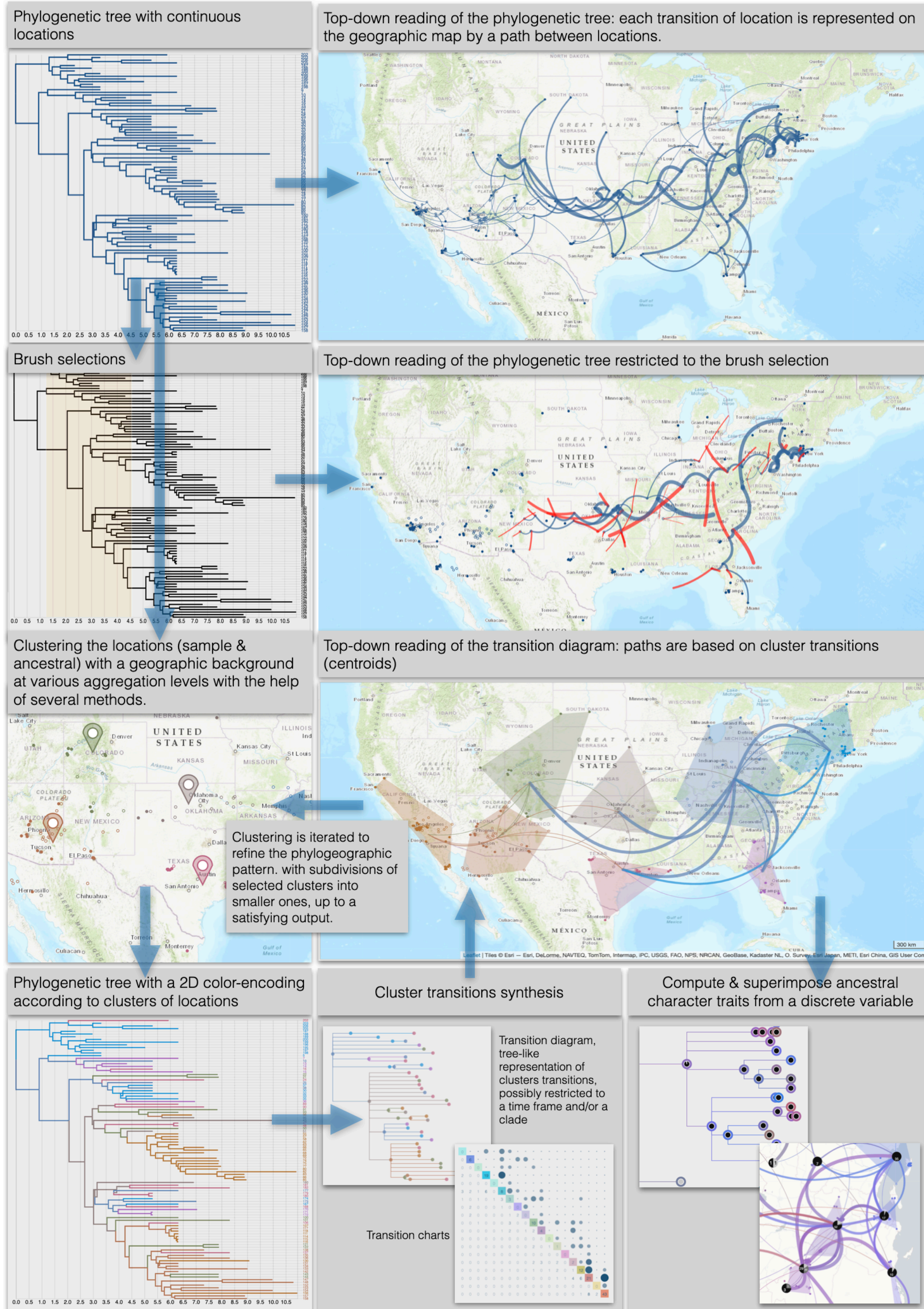


EvoLaps Manual



EvoLaps aims at visualizing spatio-temporal spread of epidemics, from phylogenetic trees associated with continuous localities (computed ancestral latitude/longitude pairs associated with computed ancestral sequences of the tree from sampled latitude/longitude pairs associated with sampled sequences).

Changes of localities, in a "Top-Down" reading of the tree (from the root to its leaves), are represented on a cartographic background using arcs between them. The bundle of arcs is a phylogeographic scenario.

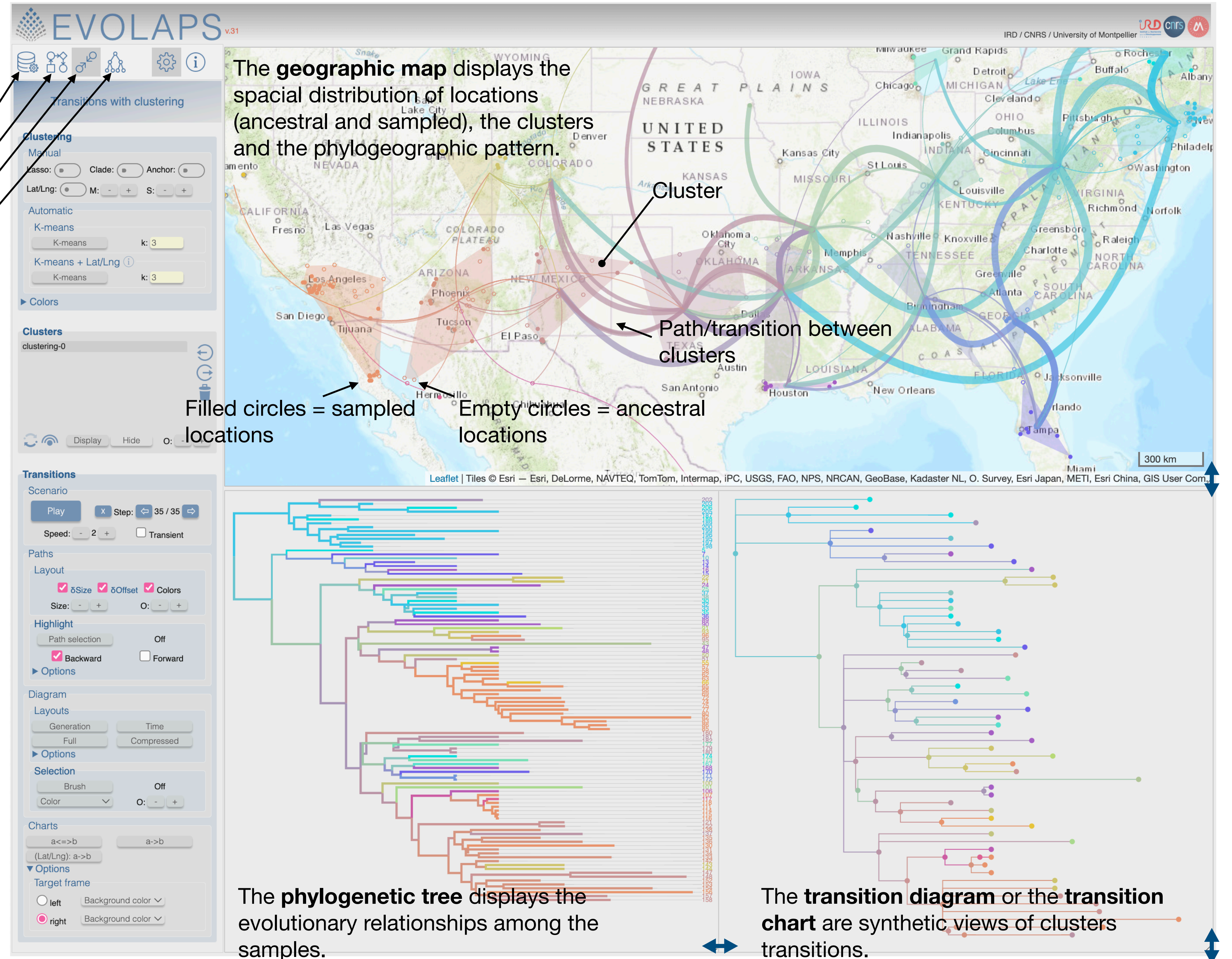
A raw reading of these localities produces complex scenarios, and EvoLaps help to analyze them with the help brush selections, and by discretizing the continuous localities into clusters of regions using different methods/tools of clustering (Lasso, Anchor, Clade selections, Dynamic latitude/longitude grid, K-means).

The dynamic EvoLaps cluster definition is iterative, allowing more or less detailed phylogeographic scenarios. For instance, an analysis starts with a small number of large clusters, and selected clusters can be subdivided afterwards, in order to detail a particular phase of the scenario.

Moreover, EvoLaps enables the computation of ancestral character states from a discrete variable, and their superimposition on the phylogeographic scenario.

EvoLaps interface

Four toolboxes:
Data Import/Export
Transition analysis without clustering
Transition analysis with clustering
Computation of ancestral character states



Data toolbox

The screenshot shows the 'Data toolbox' interface. At the top, there are icons for data, phylogenetic trees, and settings. Below this, there are three main sections: 'Import - NEXUS format', 'EvoLaps', and 'Demo'. Each section has a 'Submit' button and a file selection area. The 'Demo' section lists four examples with their respective 'Submit' buttons and file icons. At the bottom, there is an 'Output' section with 'Save analysis' and 'Save as SVG' buttons.

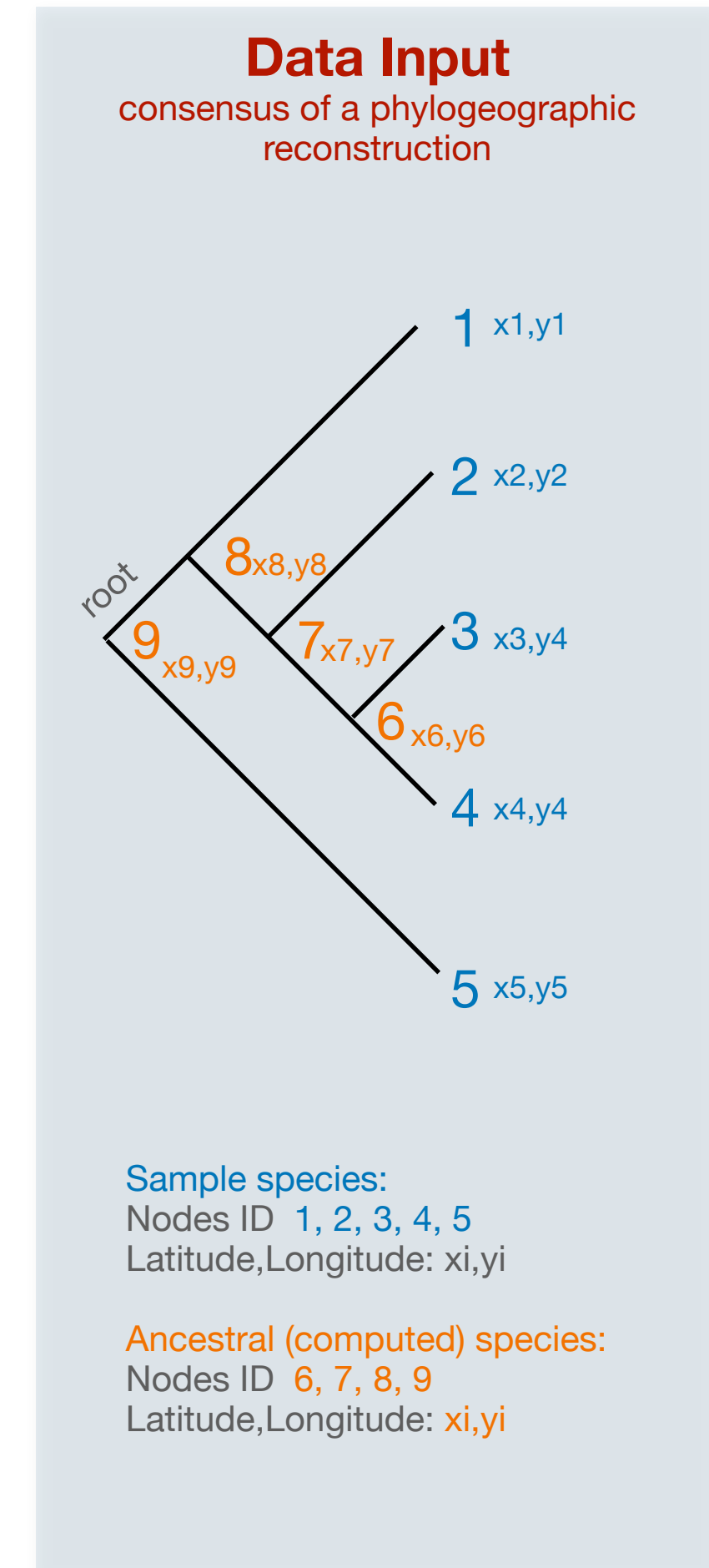
Import of data. The first step is the importation of data from a third-party software that generates a **consensus** of a phylogeographic tree. Input data (NEXUS format) must contain a rooted tree with sample and ancestral (consensus) lat/Ing coordinates for the tips (samples) and the internal nodes of the tree (ancestral species), respectively.

The variable names saving the coordinates must be as following :

Samples: `location1 location2`
 ↑ ↑
 Latitude Longitude

Ancestral(*): `location1_median location2_median`
 ↑ ↑
 Latitude Longitude

(*) if the « location1_median » and « location2_median » variables are not found during the import of the NEXUS file, the parsing process will search for « location1 » and « location2 » variables for ancestral nodes



Transition analysis without clustering

- Size and Opacity of locations (ancestral & sampled)

- Tiles layers

- Tree info

- Tree layout

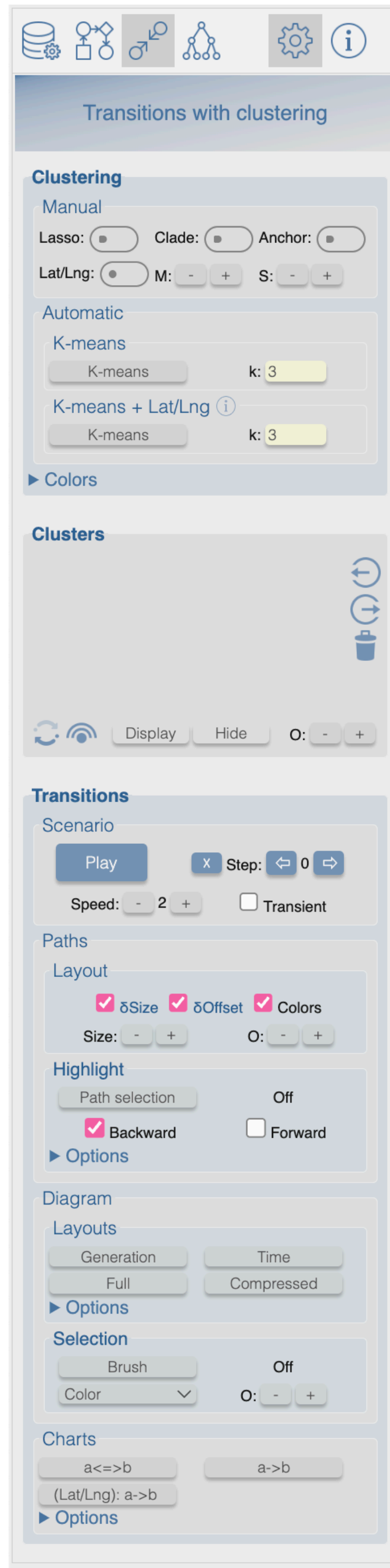
- Mode Radial/Cartesian
- Vertical expansion
- Horizontal expansion
- Edges width
- Leaf label font (size)
- Scale

- Transitions

- Drag & Drop the brush
- Resize the brush

- 'Play': the phylogeographic scenario is displayed smoothly at the given speed, thanks to a brush covering the tree. This brush is resizable and moveable.
- Speed of the animation
- Brush on/off: Click & Drag the phylogenetic tree. The brush is moveable and resizable, the phylogeographic scenario is restricted to the brush selection
- Background color and opacity of the brush
- Brush animation
 - set a number of steps and clic the 'Run' button.
 - 'Sliding' ON: a brush width is computed and the brush is moved from the tree root to its leaves, the phylogeographic scenario is displayed step by step
 - 'Sliding' OFF: an initial brush width is computed, then for each step the brush width is increased of this initial width up to the leaves
 - 'Full height' ON: the brush height covers the tree height
 - 'Save map' & 'Save tree', if checked, a SVG export of the map and/or the tree is done a each step
- paths layout 'δSize' & 'δOffset': width and curvature of paths linked to the depth of the transition in the scenario
- 'Size' and 'O': width and opacity of paths. Relative values
- 'Colors'
 - 'Full bound': Color of full bound . If a node and one of its children are inside the brush selection, the corresponding path is said « full ».
 - 'Inbound': Inbound color. A path is said « Inbound » if the corresponding children node is inside the brush selection, but not the corresponding parent node
 - 'Outbound': Outbound color. A path is said « Outbound » if the corresponding parent node from the tree is inside the brush selection but not the corresponding children node
 - 'Inter': Inter bound color. A path is said « Inter » if neither the corresponding parent node nor the corresponding children are inside the brush selection
- 'Highlight', path selection on the geographic map highlights backward/forward paths

Transition analysis with clustering



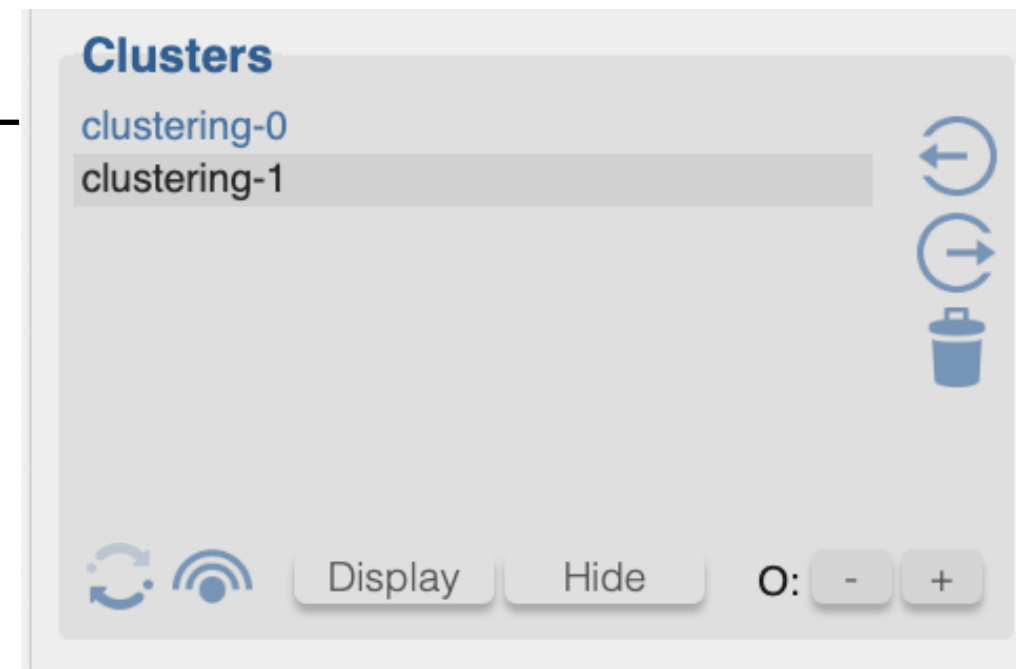
- **Lasso clustering** from the geographical map, click and drag the mouse around a set of locations to define a new cluster. Color is set automatically with the cluster centroid if 'Color scale' is checked, from the color picker otherwise
- **Clade clustering** select a node from the phylogenetic tree. Color is set automatically with the cluster centroid if 'Color scale' is checked, from the current color picker otherwise
- **Anchor clustering** click the geographic map to deposit an anchor, anchor can be moved, or deleted (ctrl-click). Color is set automatically with the cluster centroid if 'Color scale' is checked, from the the color picker otherwise. Moving an anchor do not change its color
- **Lat/Lng clustering** grid of parallel/meridian bounds. M: mesh density of Lat/Lng grid, S: bounds size. Bounds can be dragged and dropped to produce a more accurate space division.

- **K-means** algorithm for clustering. Set the desired number of clusters (3 by default) and click the « K-means » button. Click it again until a satisfying output.
- **K-means + Lat/Lng** K-means output set minimum and maximum of latitude/longitude bounds.

• clustering —

• clusters colors —

• clusters list —



- submit ans save the current clusters
- re-submit a cluster list
- delete a cluster list
- reset clustering, display locations without cluster, display/hidden clusters, 'o': clusters opacity

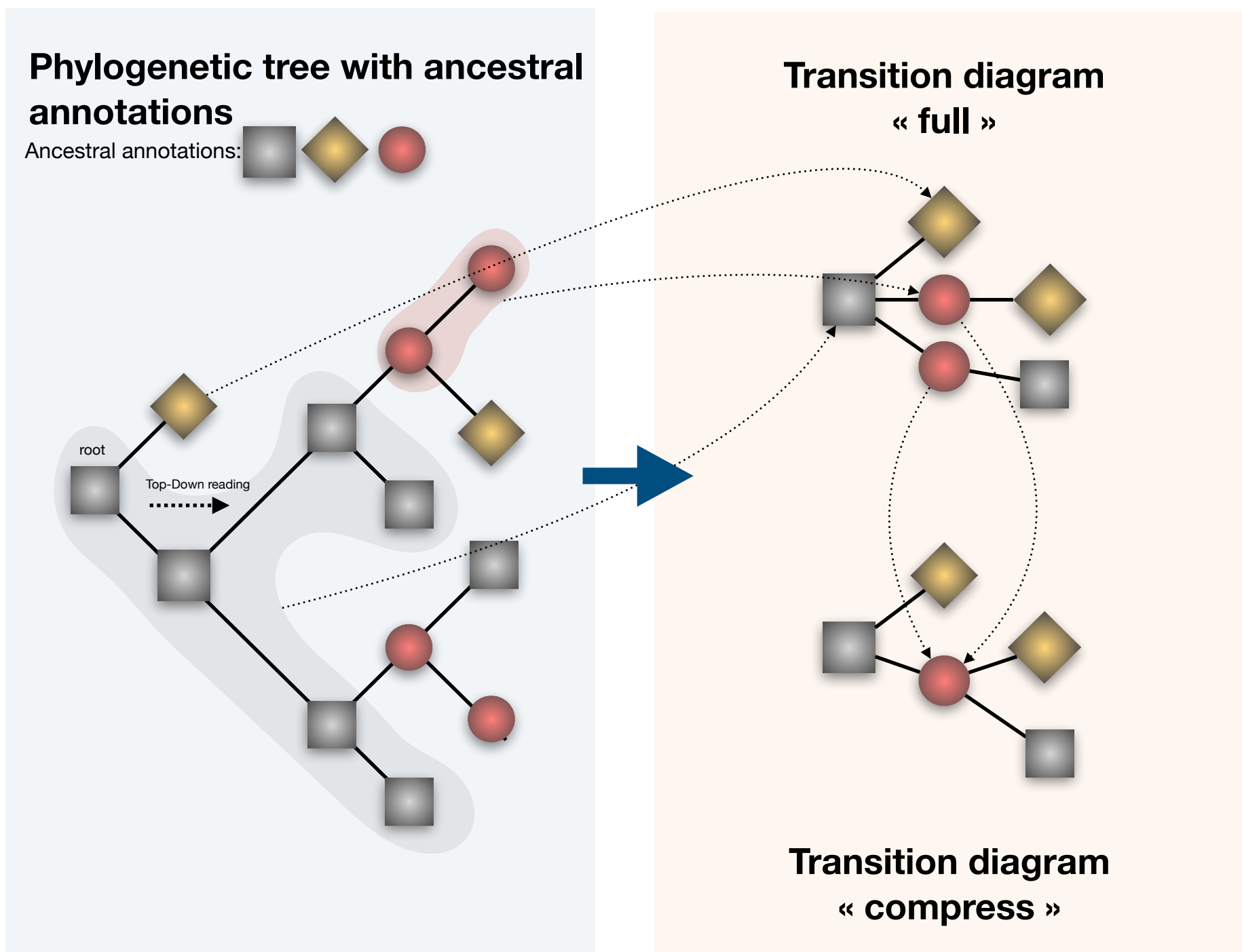
- Color picker
- Colors from clusters. Clic a cluster color and the current color (color picker) will be updated with this cluster color. 'Command-clic' a cluster color to update it with the current color (color picker).
- 2D color matrices

• transitions

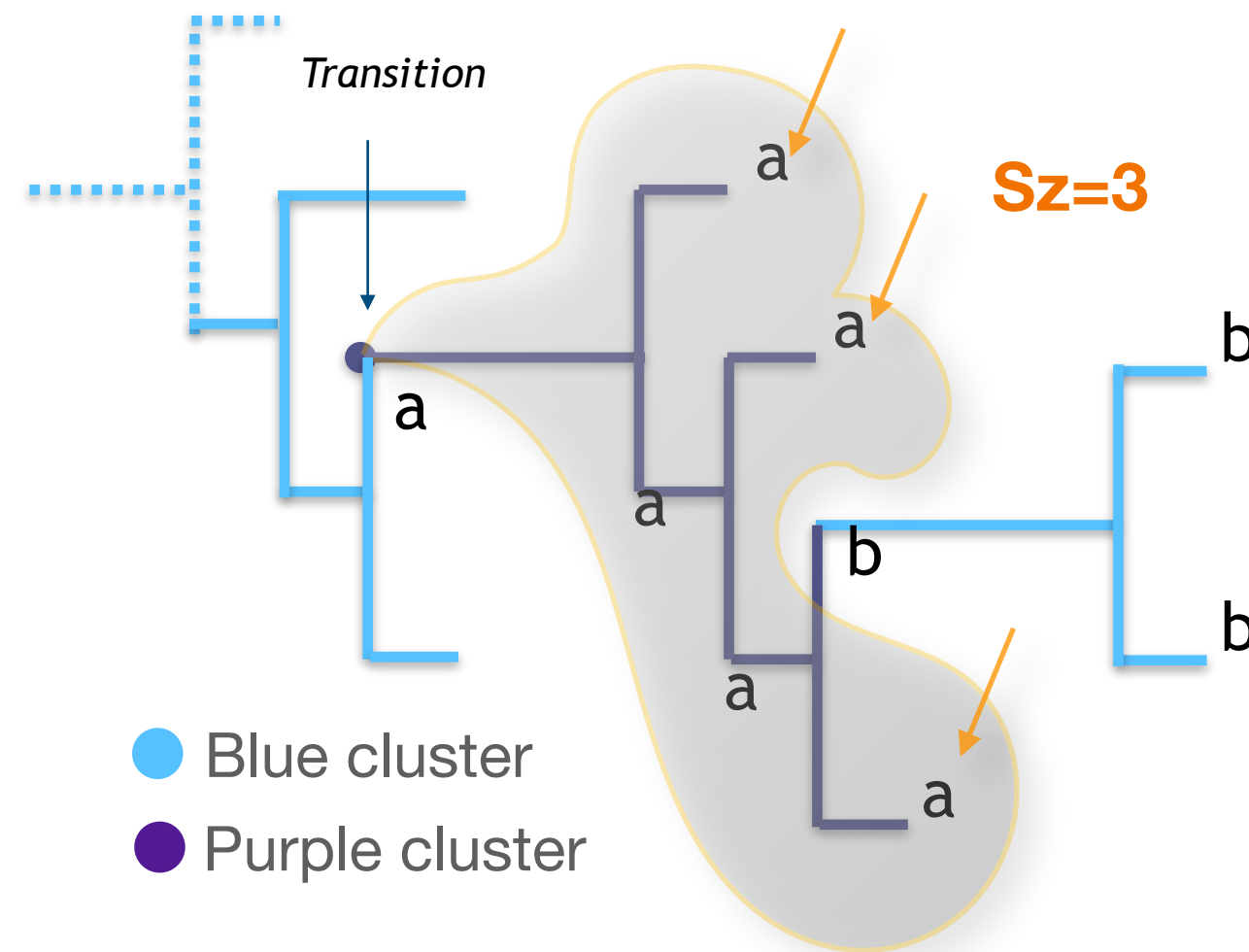
- 'Play': animation of the phylogeographic pattern
- 'X': reset the phylogeographic scenario
- '<=' & '=>': display the phylogeographic pattern step by step
- speed of the animation
- 'Transient': display only the current path
- 'δSize' & 'δOffset': width and curvature of paths linked to the step/time of the transition in the scenario
- 'Colors': paths colors linked to the cluster origin, otherwise or all paths have the same color (the current color define by the color picker)
- 'Size' and 'O': width and opacity of paths. Relative values
- 'Highlight', path selection on the geographic map highlights backward/forward paths
- Diagram layout
- Diagram layout options: vertical/horizontal expansion, edges width, node size, opacity, Sz criteria (node size proportional), Sz value, font
- Brush: Clic & Drag the transition diagram to define a selection brush. The brush is moveable and resizable, the phylogeographic scenario is restricted to the selection
- background color and opacity of the brush
- Transition chart
- Transition option: target frame

Transition diagram & chart

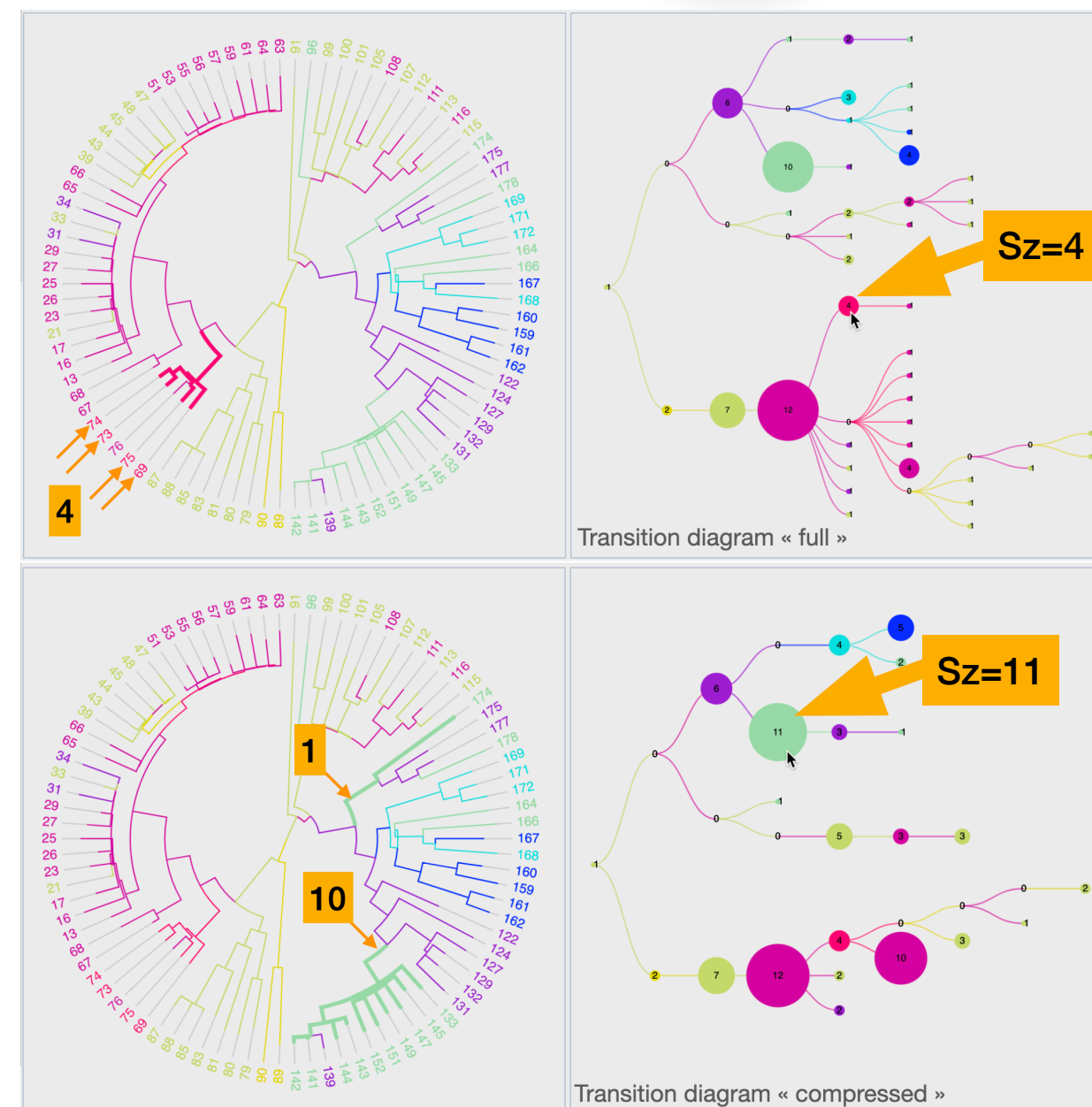
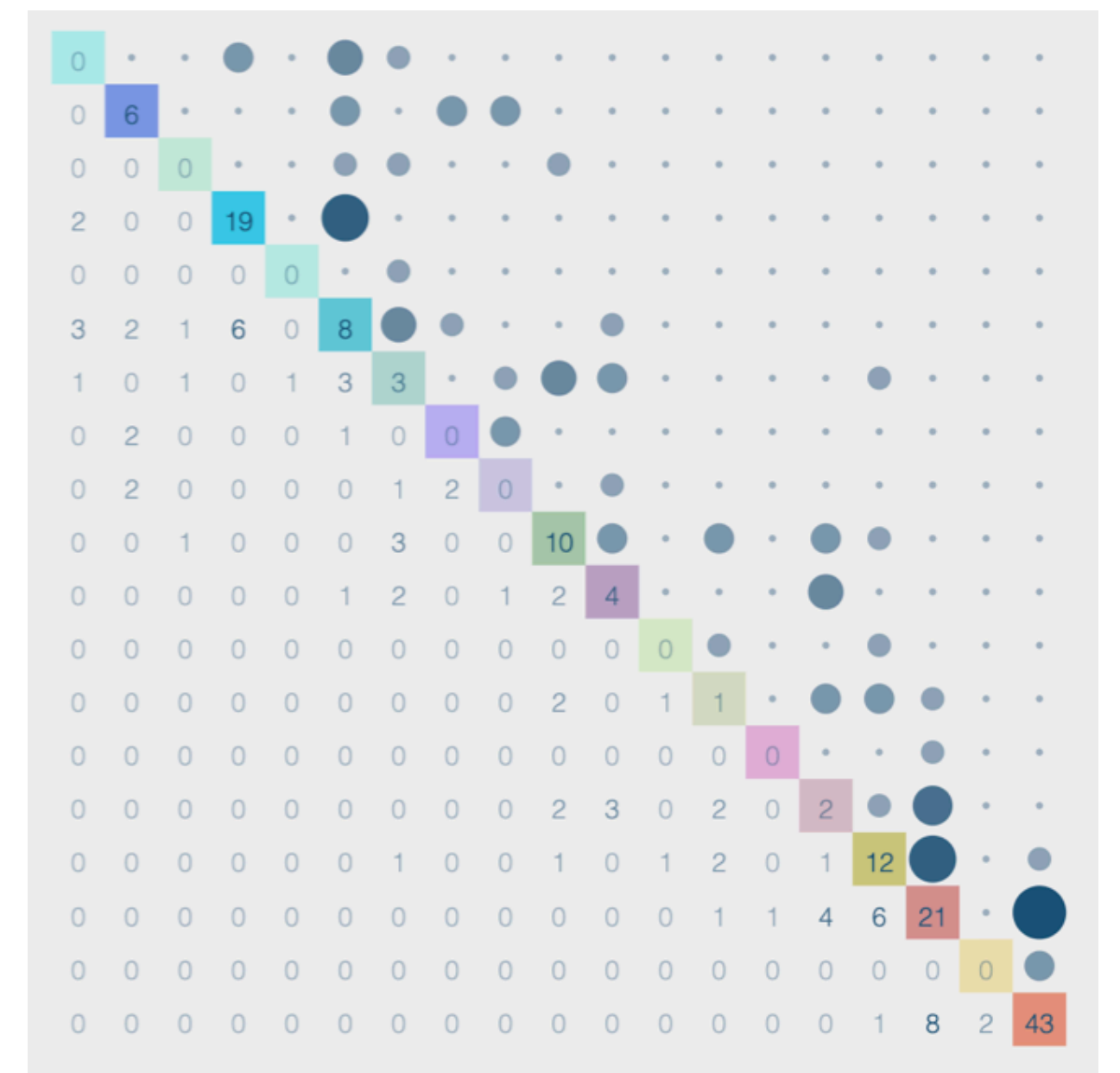
A transition is defined as an inferred change of geographical cluster between subsequent nodes of the tree in a top-down reading (from the tree root to its tips in a recursive process). The transition diagram is a multi-furcating tree-like representation, summarizing all transition suites. It gives a synthetic view of a phylogeographic pattern without the geographical constraints. A default diagram starts with a node corresponding to the ancestral root state i . A node is inserted in the transition diagram when a cluster transition $i \rightarrow j$ is observed until the tips are reached. A compressed version of the diagram is available by collapsing identical transitions having the same ancestor in the default version. The diagram is then read from its root to its tips: each transition is projected on the geographic map as a path between clusters



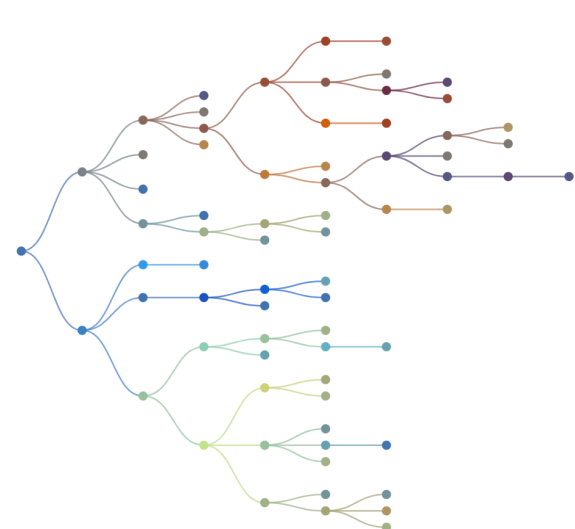
S_z is the number of descendants being in the same cluster along the path from a node of phylogenetic tree to its tips. In case of a compressed version of the transition diagram, S_z values are simply added.



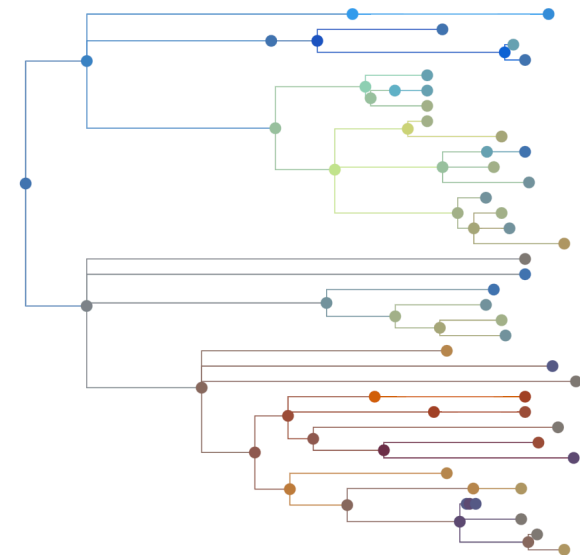
The transition chart counts the number of transitions between clusters



Layout « Generation »



Layout Time (branch lengths)



Computation of ancestral character states toolbox

Compute ancestral character states from a discrete variable (ecology, epidemiology,...) and superimpose the output to an existing phylogeographic scenario

The interface is titled 'Compute & display ancestral character traits from a discrete variable'. It features a top navigation bar with icons for database, phylogenetic tree, pie chart, and settings. The main section is divided into three parts: 1. 'Compute ancestral character states' with a file selection dropdown (currently 'Aucun fichier choisi'), model selection (radio buttons for 'F81' and 'Joint'), and prior selection (radio buttons for 'Equiprob.' and 'Tree freq.'). A 'Submit' button is located below. 2. 'Display ancestral characters' with a text input field for 'Ancestral character set(s)'. 3. 'Options' section with a 'Threshold' dropdown (set to 'M'), a 'Color scale' dropdown (set to 'grey single hue'), and an 'Opacity' control with minus and plus buttons. The 'Target' section at the bottom has three checkboxes: 'Geographic map' (unchecked), 'Phylogenetic tree' (unchecked), and 'Transition diagram' (checked). Each checkbox has a 'size' control with minus and plus buttons.

1. Select a csv file putting in regards sequence labels from the phylogenetic tree with a discrete variable. In this example the EcoType variable has 3 modalities : F, D and S
2. Select a method to compute ancestral character states, either « F81 » or « Joint »,
3. Select priors « Equiprob. » or « Tree freq. »
4. Submit , according to the number of sequences and the number of modalities of the discrete variable, computing ancestral character states may take a little while
5. Select an ancestral character set from the listbox, check on/off target(s) to display ancestral states on the geographic map, the phylogenetic tree and the transition diagram

- Treshold used to display ancestral character states (M= majority)
- Colors scale of piecharts displaying ancestral character states
- Opacity of piecharts displaying ancestral character states
- Check on/off to display piecharts of ancestral character states
- Size of piecharts displaying ancestral character states

The terminal window is titled 'DEMO1-EcoType.txt' and displays a list of sequence IDs and their corresponding EcoType modalities. The output is as follows:

```
ID, EcoType
2013Be591cp.seq, F
2009Ni105cp.seq, F
2012Gh3387cp.seq, F
2013Tg241cp.seq, F
1975CI17cp.seq, F
1975CI68cp.seq, F
1975CIacp.seq, F
1975SL1cp.seq, F
1975SL2cp.seq, F
1975SL3cp.seq, F
1975SL4cp.seq, F
1975SL5cp.seq, F
1975SL6cp.seq, F
1975SL7cp.seq, F
1980Ni131cp.seq, F
1980Ni1cp.seq, D
1980Ni2cp.seq, F
1980Ni4cp.seq, F
1980Ni6cp.seq, F
1980Ni1cp.seq, F
1990BF1cp.seq, S
1990BF2cp.seq, S
1990CIbcp.seq, F
1990Ma1cp.seq, S
1990Ma2cp.seq, S
1991CI12cp.seq, F
1991CI13cp.seq, F
1991Gu1cp.seq, F
1992Ma3cp.seq, D
1992Ma9cp.seq, D
1994CI10cp.seq, F
```

- Ancestral character states are NOT dependent of the clustering used to establish the phylogeographic scenario.
- **Geographic map** A piechart associated to a cluster displays the number of times each modality of the discrete variable under study has a probability \geq of a threshold, for each of the nodes underlying to the cluster
- **Phylogenetic tree** For each node of the phylogenetic tree, the associated piecharts display the modalities of the discrete variable under study having a probability \geq of a threshold. The threshold is the probability of the majority for the given node or the probability of the majority minus a percentage of its value.
- **Transition diagram** A Piechart associated to a node of the transition diagram displays the number of times each modality of the discrete variable under study has a probability \geq of a threshold, for each of the tree nodes and the sequences underlying the node of the transition diagram. The threshold is the probability of the majority or the probability of the majority minus a percentage of its value.

Appendix

Computing the marginal posterior probabilities of every state for each of the tree nodes (F-81-like) and the joint ancestral scenario with the maximal posterior probability

Computing the marginal posteriors probabilities (F81-like)

Probability of changes along a branch of length l is expressed as:

$$\begin{aligned} \text{PC}(i \rightarrow j / l) &= (1 - e^{-\mu l}) \pi_j \text{ if } j \neq i \\ &= e^{-\mu l} + (1 - e^{-\mu l}) \pi_i \text{ otherwise} \end{aligned}$$

With π_i equilibrium frequency (priors)

user supplied, roughly estimated from the state frequencies observed at the tree tips or equal probability

with $\mu = 1 / (1 - \sum_i \pi_i^2)$ (normalization factor)

Computing the marginal posteriors probabilities is done in four steps:

- 1** tree rescaling. maximize the tree likelihood to find the scaling factor with iterative calls to a Pruning (bottom-up) with a gold number or Brent search. The tree is then rescaling with the found scaling factor
- 2** bottom-up computation of 'down' conditional likelihoods (post-order recursion using the pruning algorithm)
- 3** top-down computation of 'up' conditional likelihoods (pre-order recursion)
- 4** computation of the states marginal posterior probabilities knowing the 'down' and 'up' conditional likelihoods

Computing the marginal posterior probabilities (F81-like)

- 1 tree rescaling. maximize the tree likelihood to find the scaling factor with iterative calls to a Pruning (bottom-up) with a gold number or Brent search. The tree is then rescaling with the found scaling factor. G, D = Left and Right descendant respectively

$$L(T) = \sum_i \pi_i L_{\text{down}}(T/i)$$

$$L_{\text{down}}(T/i) = \left[\sum_j PC(i \rightarrow j/l_g) L_{\text{down}}(G/j) \right] * \left[\sum_j PC(i \rightarrow j/l_d) L_{\text{down}}(D/j) \right]$$

$$L_{\text{down}}(F/i) = 1 \text{ if the tip } F \text{ is annotated with } i, 0 \text{ otherwise}$$

- 2 bottom-up computation of 'down' conditional likelihoods (post-order recursion using the pruning algorithm)

- 3 top-down computation of 'up' conditional likelihoods (pre-order recursion)

The 'up' likelihood of G is calculated from the 'down' likelihood of D and the 'up' likelihood of their common father, and vice versa.

$$L_{\text{up}}(G/i) = \left[\sum_j PC(i \rightarrow j/l_g) L_{\text{up}}(T/j) \left[\sum_k PC(j \rightarrow k/l_d) L_{\text{down}}(D/k) \right] \right]$$

$$L_{\text{up}}(D/i) = \left[\sum_j PC(i \rightarrow j/l_d) L_{\text{up}}(T/j) \left[\sum_k PC(j \rightarrow k/l_g) L_{\text{down}}(G/k) \right] \right]$$

$$L_{\text{up}}(T/j) = 1 \text{ si } T \text{ is the whole tree}$$

- 4 computation of the states marginal posterior probabilities knowing the 'down' and 'up' conditional likelihoods

$$\text{Post}(N, i) = \pi_i L_{\text{down}}(N/i) L_{\text{up}}(N/i) / L(T)$$

$$L(T) = \sum_i \pi_i L_{\text{down}}(N/i) L_{\text{up}}(N/i)$$

The pruning algorithm (Felsenstein 1981)

$$L(T) = \sum_i \pi_i L_{\text{down}}(T/i)$$

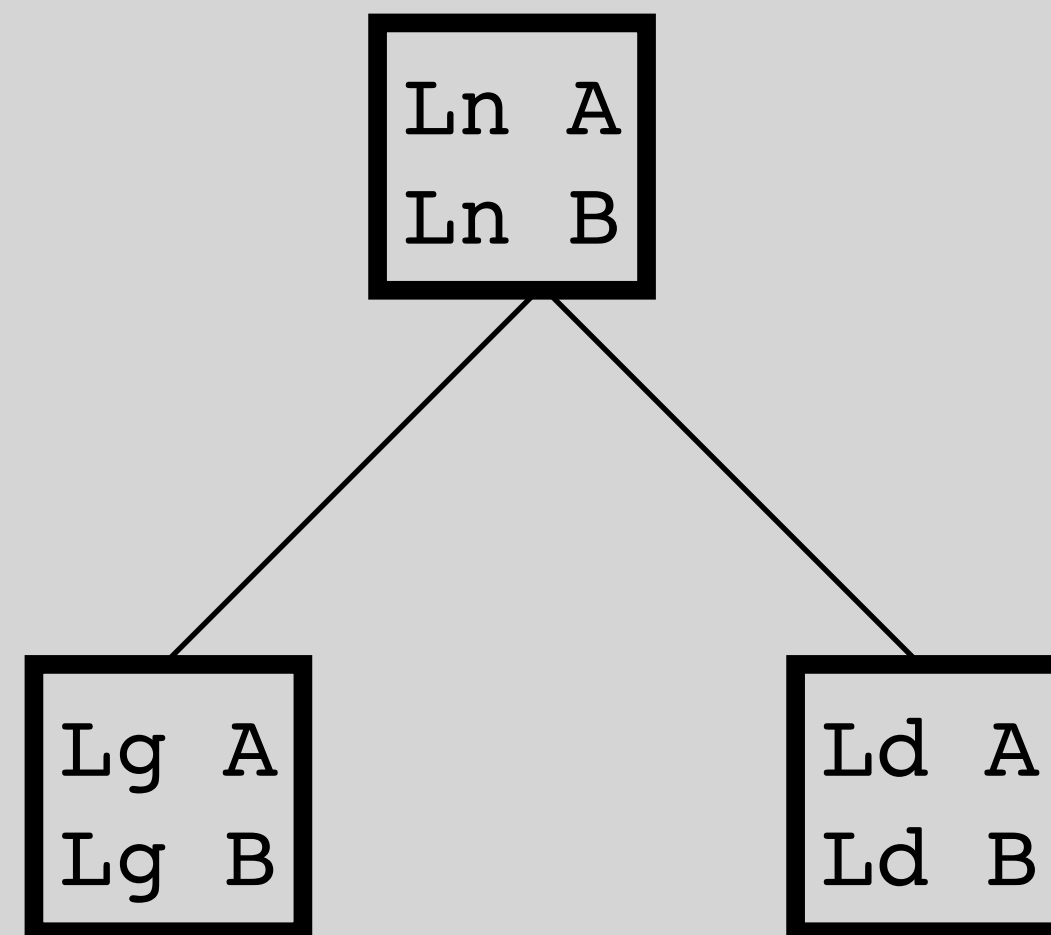
$$L_{\text{down}}(T/i) = \left[\sum_j PC(i \rightarrow j / l_g) L_{\text{down}}(G/j) \right] * \left[\sum_j PC(i \rightarrow j / l_d) L_{\text{down}}(D/j) \right]$$

$L_{\text{down}}(F/i) = 1$ if the tip F is annotated with i, 0 otherwise

$$L = \pi_A * L_nA + \pi_B * L_nB$$

States A and B
G left descendant
D right descendant

Conditional
Likelihood Knowing A,
Left descendant



« post order »

$$L_nA = (L_gA * PC(A \rightarrow A) + L_gB * PC(A \rightarrow B)) * (L_dA * PC(A \rightarrow A) + L_dB * PC(A \rightarrow B))$$

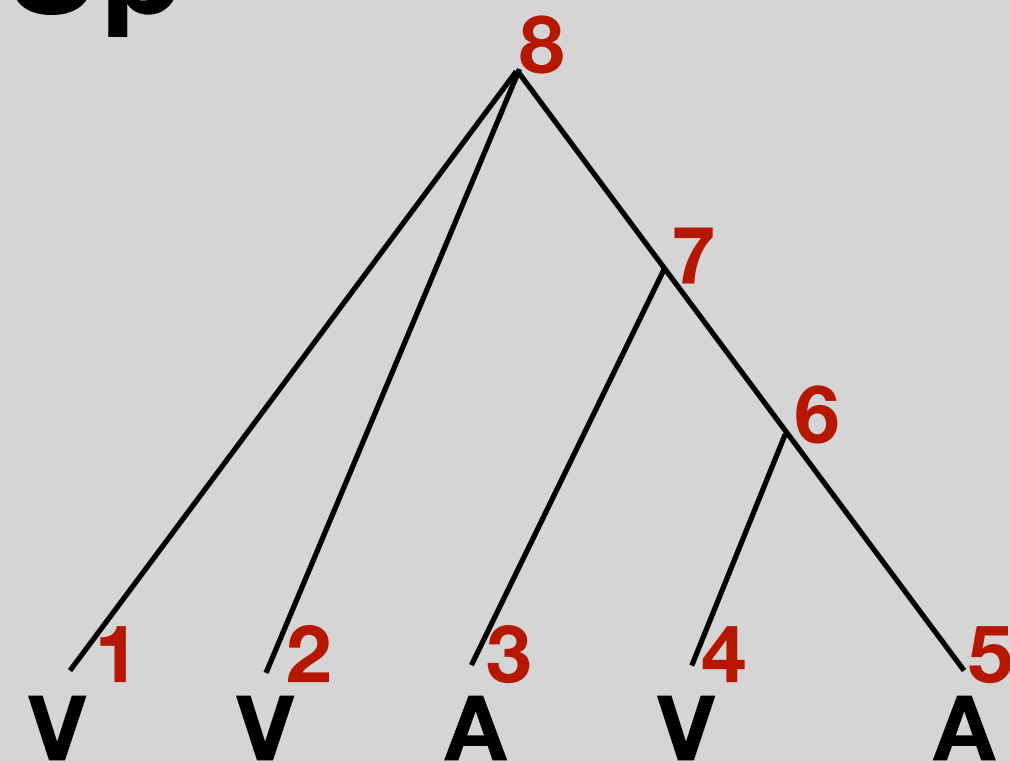
$$L_nB = (L_gA * PC(B \rightarrow A) + L_gB * PC(B \rightarrow B)) * (L_dA * PC(B \rightarrow A) + L_dB * PC(B \rightarrow B))$$

Computing the joint scenario, example

(Yang, 1995) (Pupko, 2000)

Tree scaling, then : Up + Down

Up



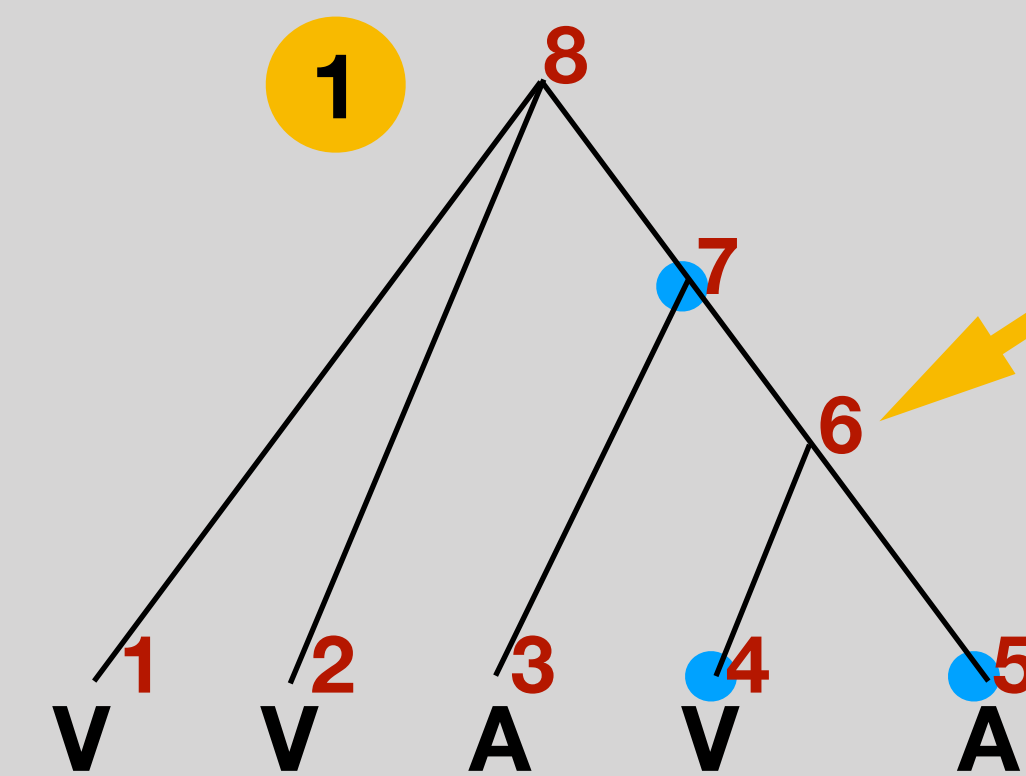
Priors

$$\pi_A = 0.4 \text{ (2/5)}$$

$$\pi_V = 0.6 \text{ (3/5)}$$

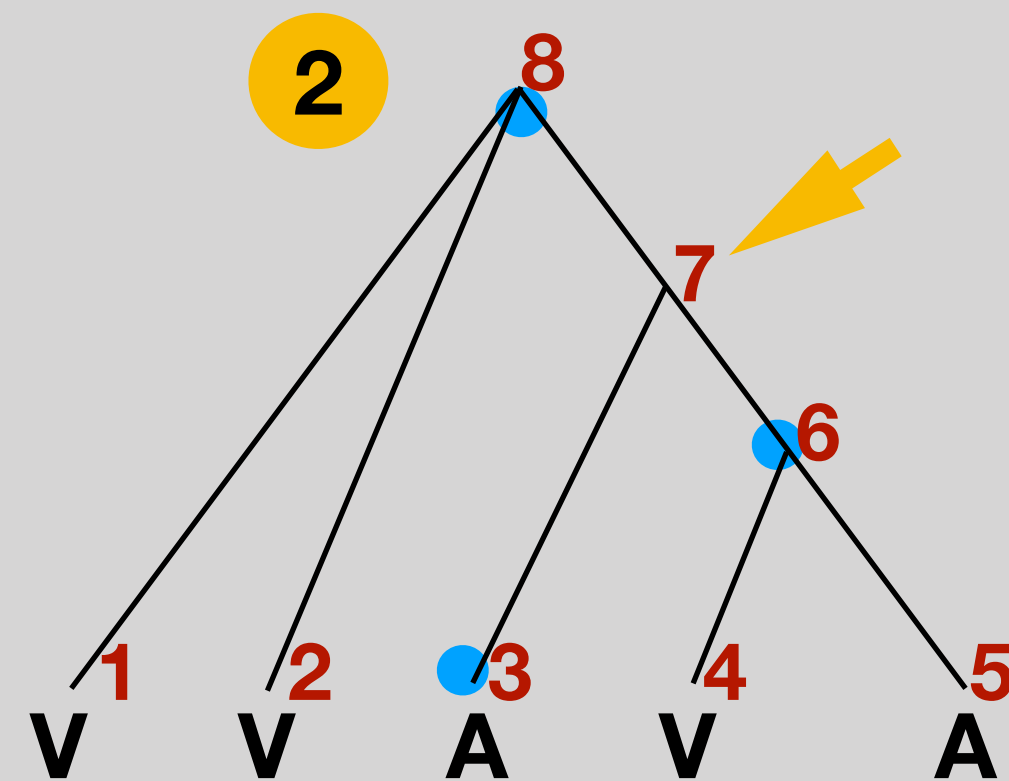
Probabilities of change

PC From	To A	V
A	0.7	0.3
V	0.45	0.55



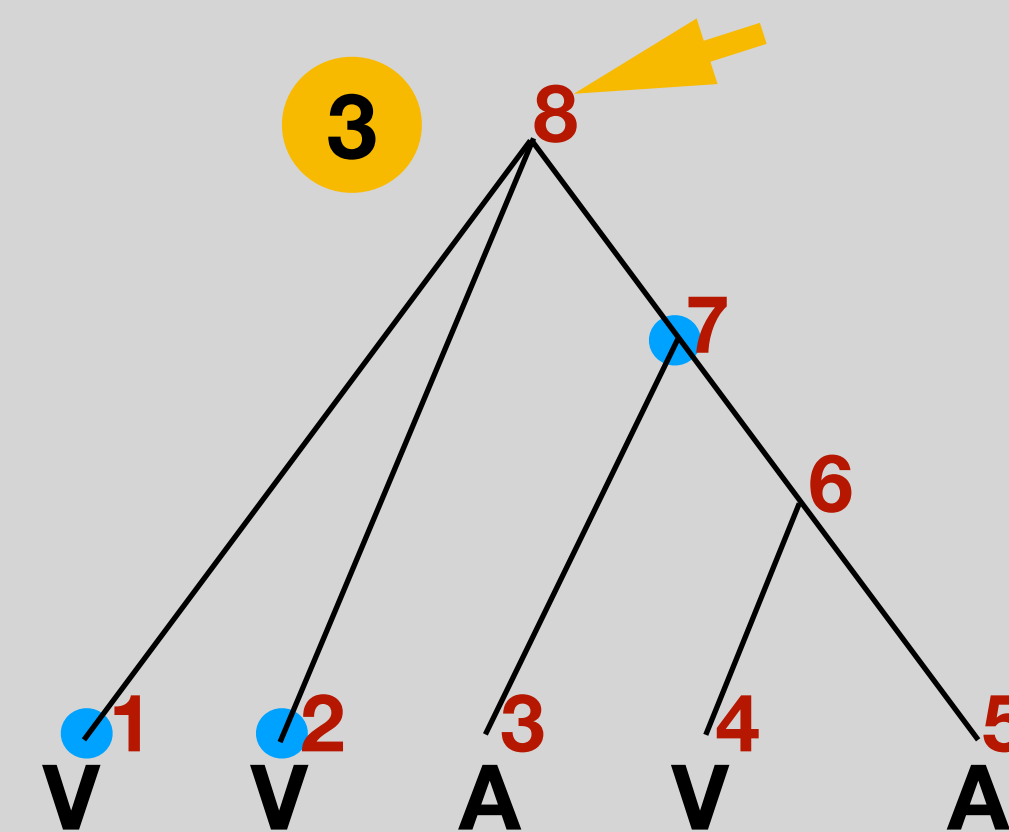
If 7 is A
 If 6 is A $L6 = 0.7 \times 0.3 \times 0.7 = 0.147 \rightarrow A$
 If 6 is V $L6 = 0.3 \times 0.55 \times 0.45 = 0.07425$

If 7 is V
 If 6 is A $L6 = 0.45 \times 0.3 \times 0.7 = 0.0945$
 If 6 is V $L6 = 0.55 \times 0.55 \times 0.45 = 0.1361 \rightarrow V$



If 8 is A
 If 7 is A $L7 = 0.7 \times 0.7 \times 0.147 = 0.072 \rightarrow A$
 If 7 is V $L7 = 0.3 \times 0.7 \times 0.1361 = 0.02856$

If 8 is V
 If 7 is A $L7 = 0.45 \times 0.7 \times 0.147 = 0.0463 \rightarrow A$
 If 7 is V $L7 = 0.55 \times 0.45 \times 0.1361 = 0.03366$



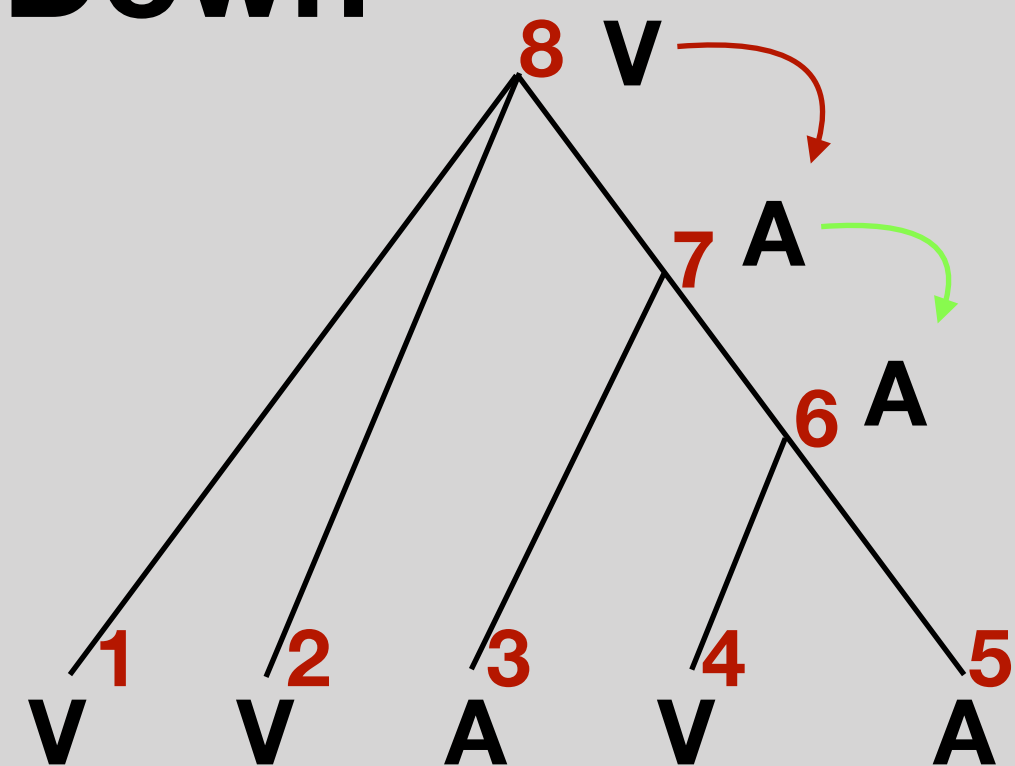
If 8 is A $L8 = 0.3 \times 0.3 \times 0.072 \times 0.4 = 0.0026$
 If 8 is V $L8 = 0.55 \times 0.55 \times 0.0463 \times 0.6 = 0.0084 \rightarrow V$

Computing the joint scenario, example

(Yang, 1995) (Pupko, 2000)

Tree scaling, then : Up + Down

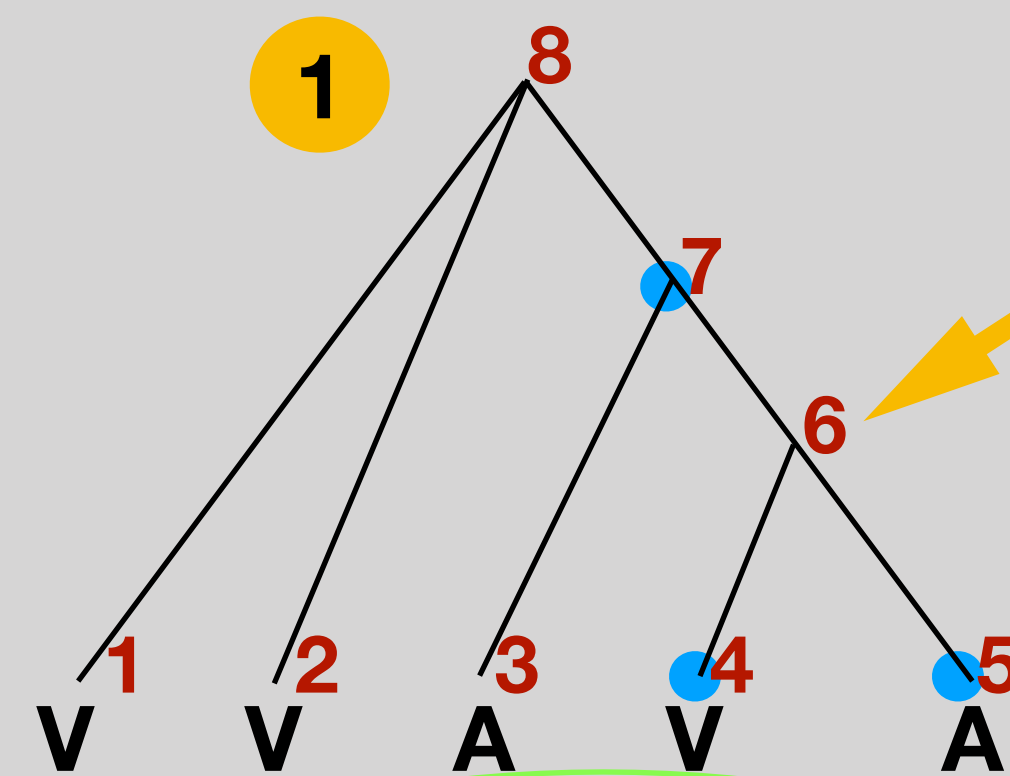
Down



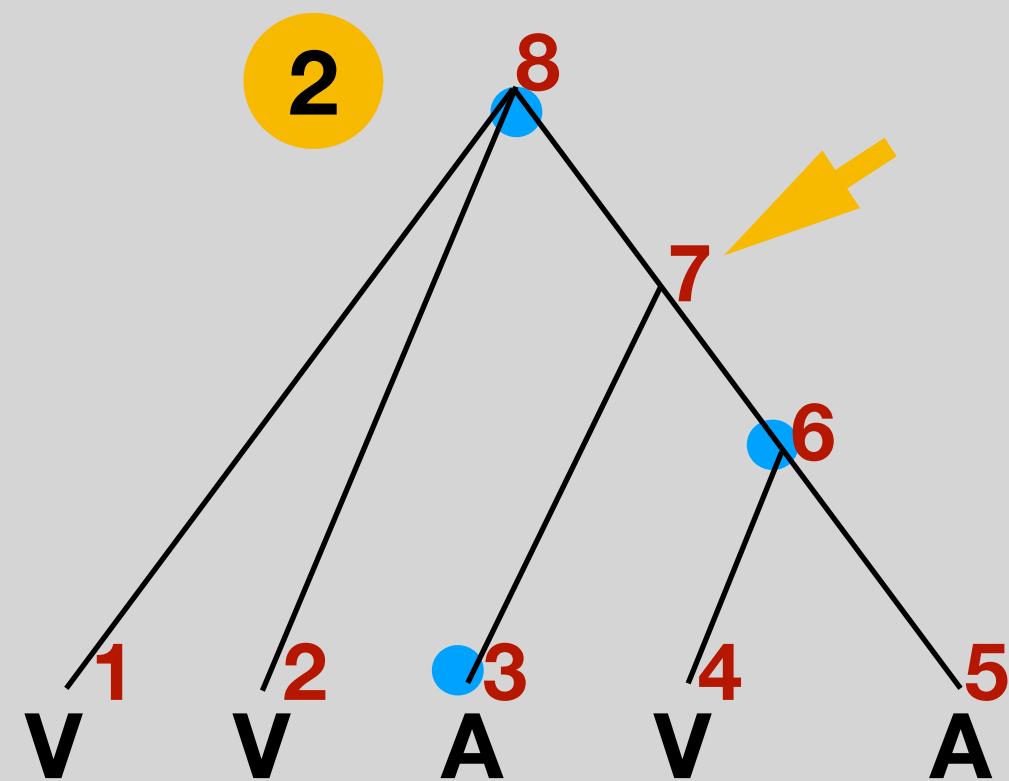
$$\pi_A = 0.4 \text{ (2/5)}$$

$$\pi_V = 0.6 \text{ (3/5)}$$

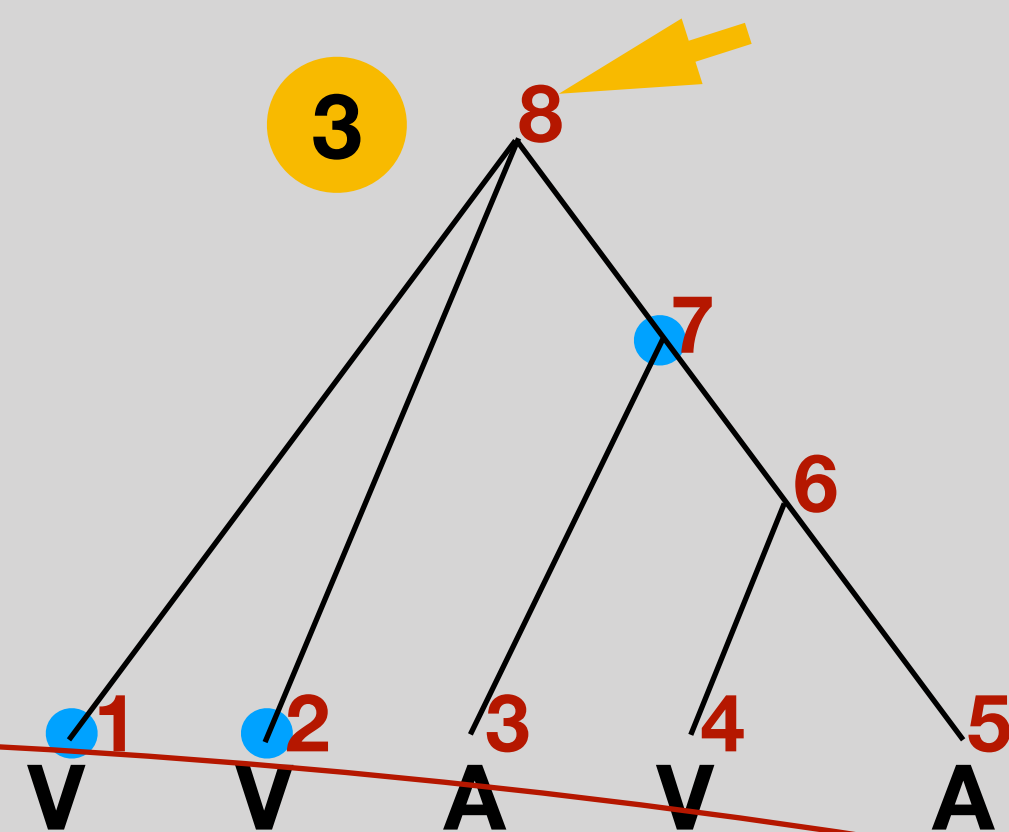
PC	Vers	A	V
De			
A		0.7	0.3
V		0.45	0.55



If 7 is A
 If 6 is A $L6 = 0.7 \times 0.3 \times 0.7 = 0.147$ → A
 If 6 is V $L6 = 0.3 \times 0.55 \times 0.45 = 0.07425$
 If 7 is V
 If 6 is A $L6 = 0.45 \times 0.3 \times 0.7 = 0.0945$
 If 6 is V $L6 = 0.55 \times 0.55 \times 0.45 = 0.1361$ → V



If 8 is A
 If 7 is A $L7 = 0.7 \times 0.7 \times 0.147 = 0.072$ → A
 If 7 is V $L7 = 0.3 \times 0.7 \times 0.1361 = 0.02856$
 If 8 is V
 If 7 is A $L7 = 0.45 \times 0.7 \times 0.147 = 0.0463$ → A
 If 7 is V $L7 = 0.55 \times 0.45 \times 0.1361 = 0.03366$



If 8 is A $L8 = 0.3 \times 0.3 \times 0.072 \times 0.4 = 0.0026$
 If 8 is V $L8 = 0.55 \times 0.55 \times 0.0463 \times 0.6 = 0.0084$ → V