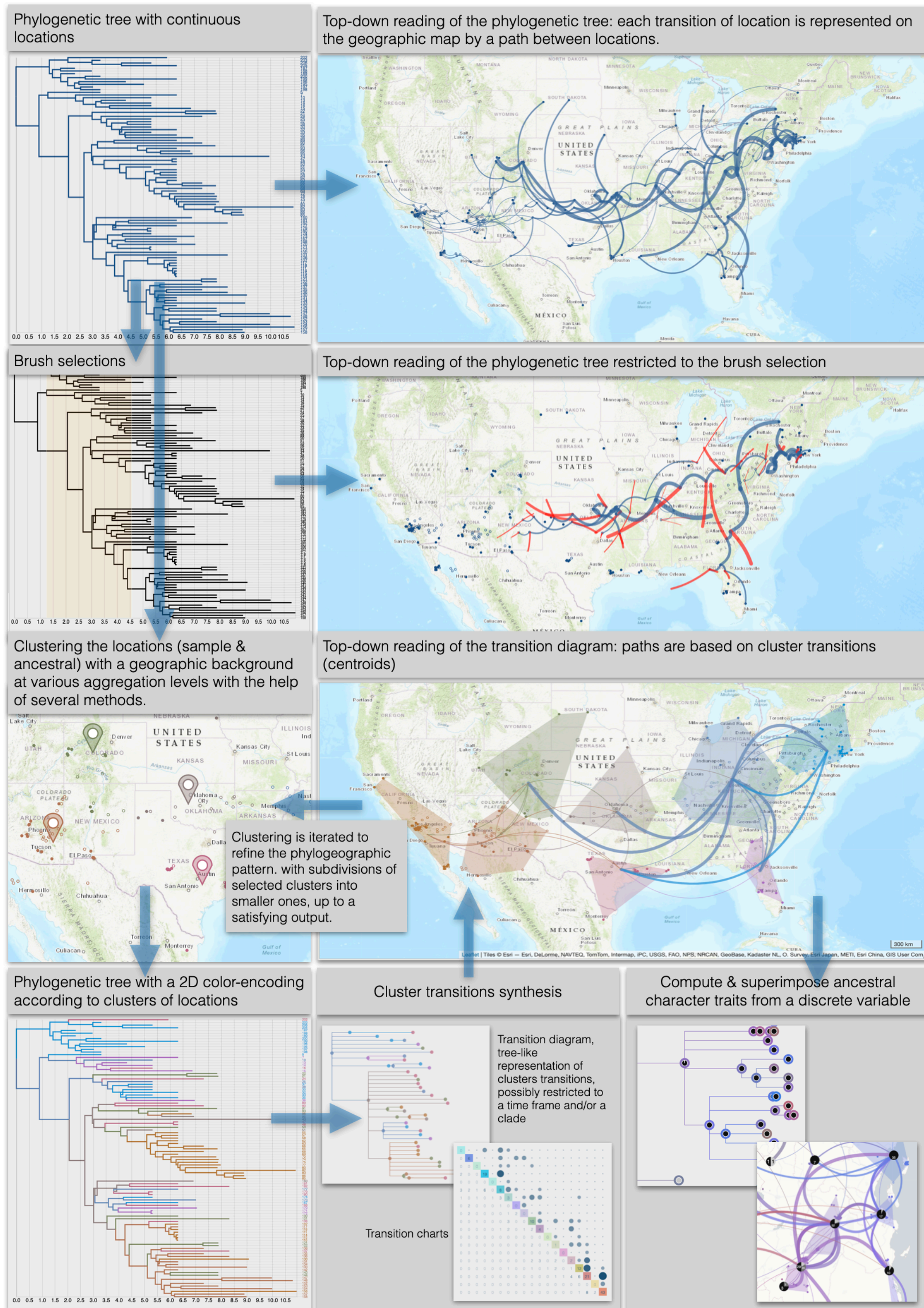


# EvoLaps Manual



EvoLaps aims at visualizing spatio-temporal spread of pathogens, from annotated Maximum Clade Credibility (MCC) tree obtained through a BEAST continuous phylogeographic inference. Changes of localities (transitions), in a "Top-Down" reading of the tree (from the root to its leaves), are represented on a cartographic background using paths between them. The bundle of paths is a phylogeographic scenario.

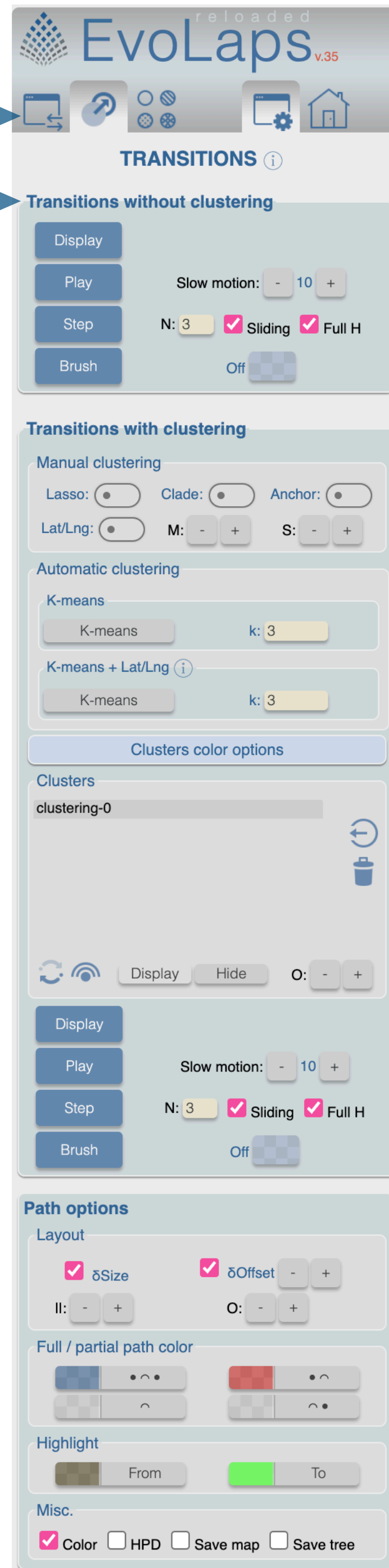
A raw reading of these transitions produces complex scenarios, and EvoLaps helps to analyze them with brush selection, animation, highlight processes, but also with the possibility of discretizing the continuous localities into spatial clusters (regions) using different methods/tools (lasso, anchor, clade selection, latitude/longitude grid, K-means). The EvoLaps clustering is dynamic and iterative, allowing more or less detailed phylogeographic scenarios according to specific spatio-temporal scales. For instance, an analysis starts with a small number of large clusters, and selected clusters can be subdivided afterwards for having a higher resolution.

EvoLaps also offers tools to corroborate third party variables to the phylogeographic scenario such as the superimposition of heat maps, or the computation of ancestral character states from discrete variables.

# EvoLaps interface

toolboxes

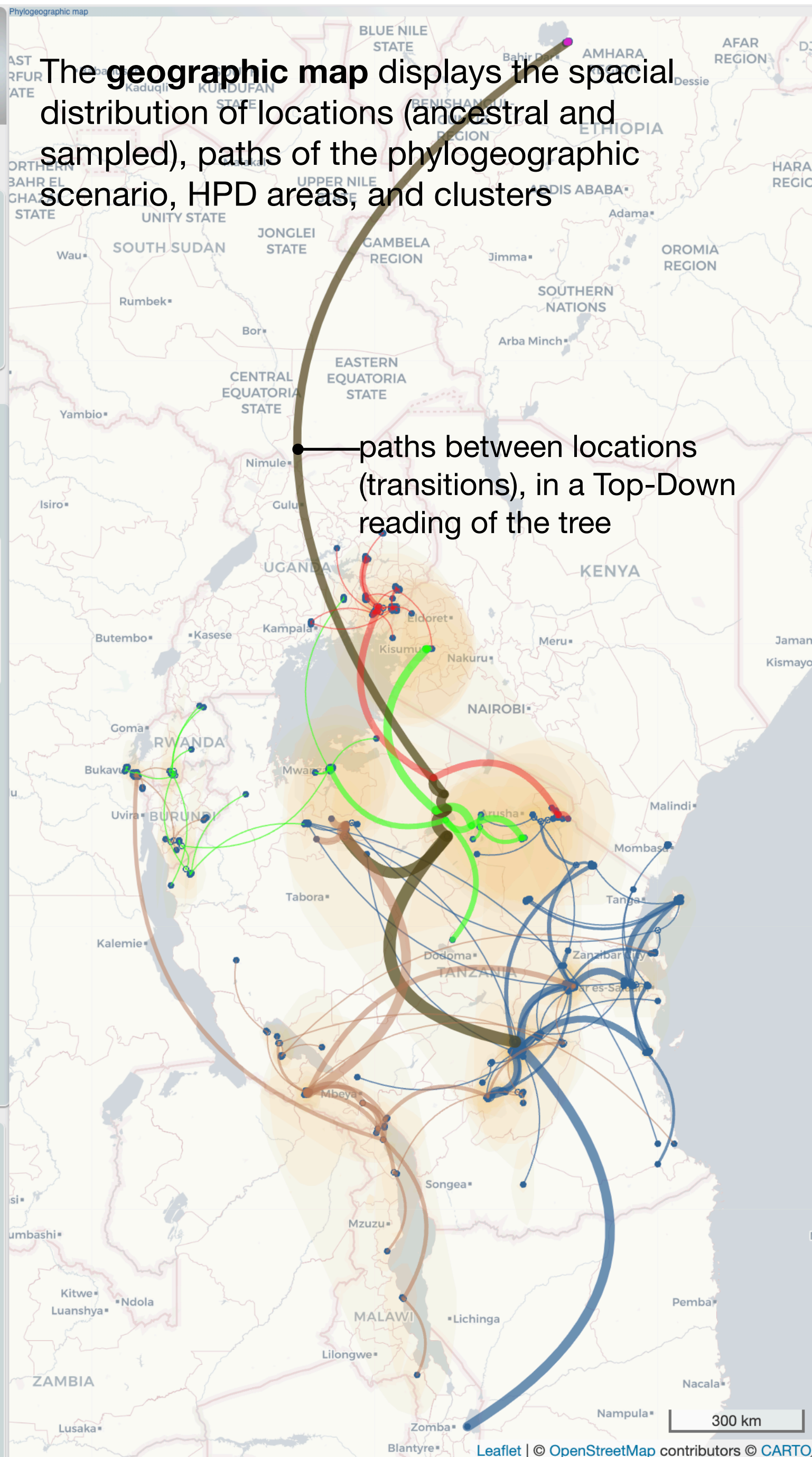
open/close boxes



The interface features several toolboxes on the left side. At the top, there are icons for home, settings, and a refresh button. Below these are two main sections: 'TRANSITIONS' and 'Path options'. The 'TRANSITIONS' section is divided into 'Transitions without clustering' and 'Transitions with clustering'. The 'without clustering' section includes buttons for 'Display', 'Play', 'Step', and 'Brush', along with a 'Slow motion' slider set to 10 and checkboxes for 'N: 3', 'Sliding', and 'Full H'. The 'with clustering' section includes 'Manual clustering' options (Lasso, Clade, Anchor, Lat/Lng, M, S) and 'Automatic clustering' options (K-means, K-means + Lat/Lng). Below these are 'Clusters color options' and a 'Clusters' list. The 'Path options' section includes 'Layout' options (deltaSize, deltaOffset, II, O), 'Full / partial path color' selection, 'Highlight' options (From, To), and 'Misc.' options (Color, HPD, Save map, Save tree).

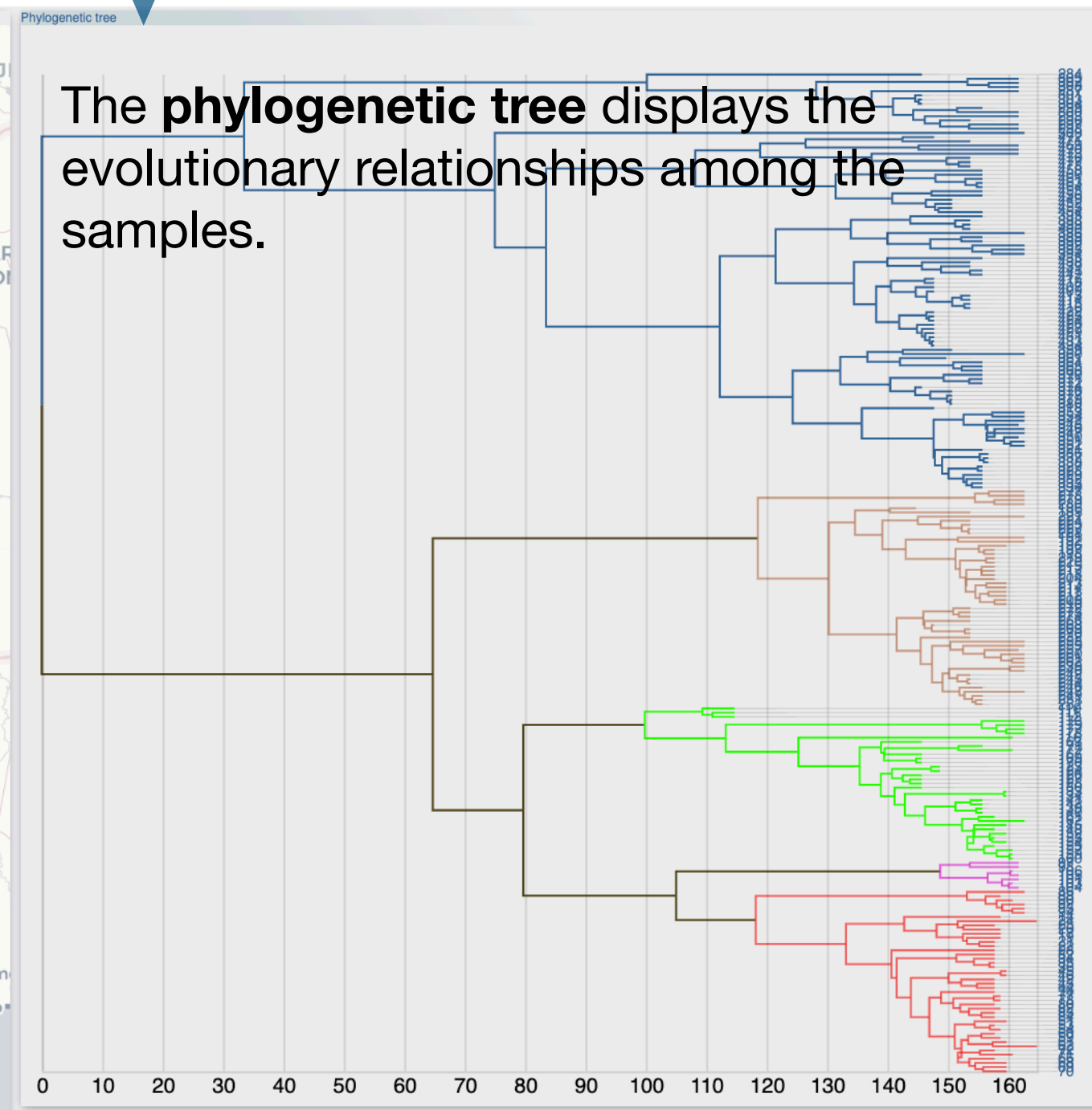
The **geographic map** displays the spatial distribution of locations (ancestral and sampled), paths of the phylogeographic scenario, HPD areas, and clusters

paths between locations (transitions), in a Top-Down reading of the tree



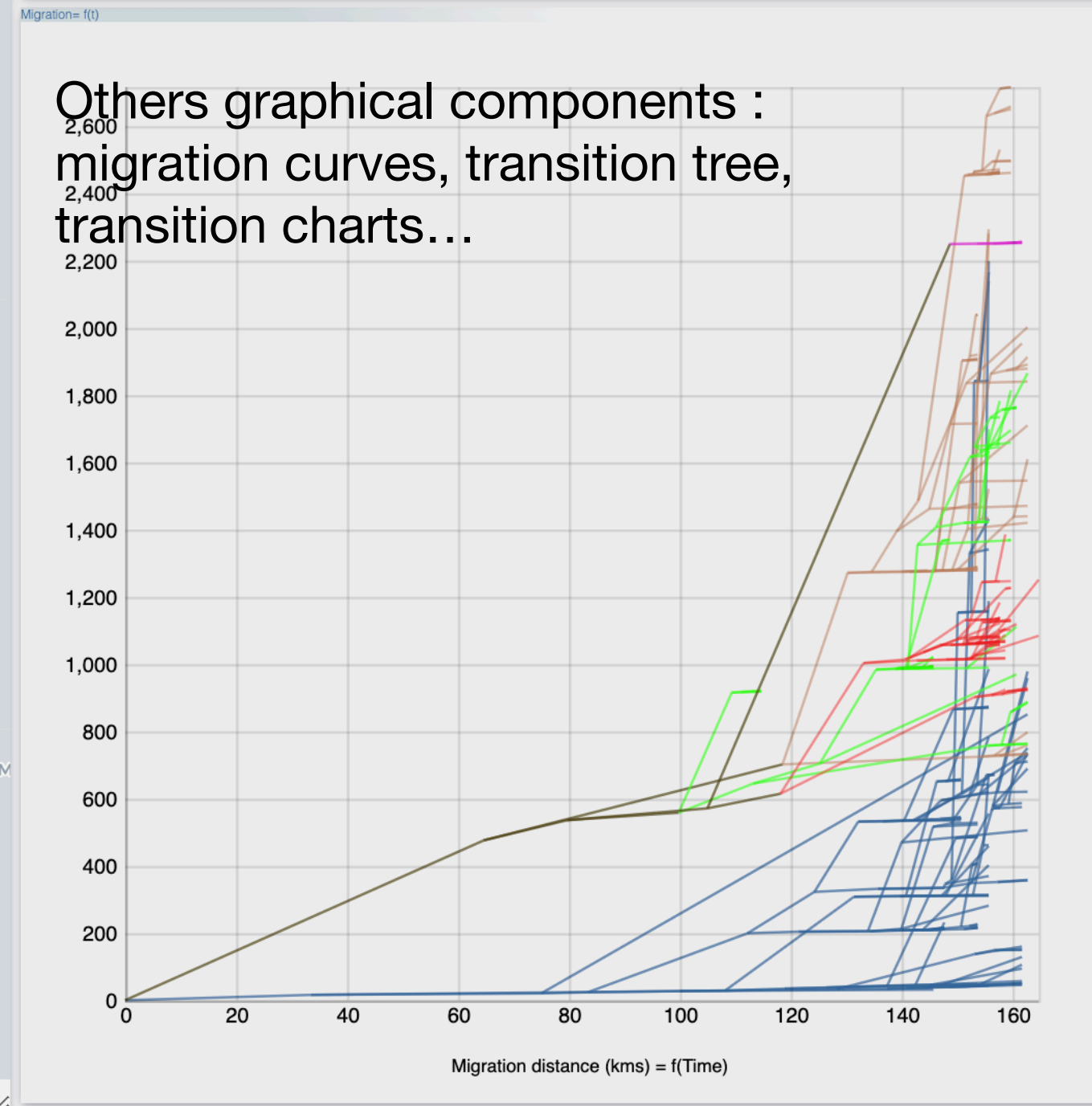
Drag & drop

The **phylogenetic tree** displays the evolutionary relationships among the samples.



Resize

Others graphical components : migration curves, transition tree, transition charts...



# EvoLaps interface

Three toolboxes:

## Data: input/output of data.

Submit a maximum clade credibility (MCC) tree obtained by continuous phylogeographic inference. Upload an evolaps analysis file. Library of data sets for demos. Save evolaps analyses or export graphic as SVG files. Preferences of the interface

## Transitions: phylogeographic scenario

as a bundle of paths between locations (transitions), with or without clustering, in a Top-Down reading of the tree. Path options.

**Edition: layouts of the geographic map and the phylogenetic tree**, and third party graphical components (transition tree, migration curve, etc.)

reloaded  
EvoLaps v.35

DATA ⓘ info tips

Import ⓘ  
Choisir un fichier Aucun fichier choisi  
Submit

Upload ⓘ  
Choisir un fichier Aucun fichier choisi  
Submit

Save/Export ⓘ  
Save Save as SVG ▾

Demo ⓘ  
Rice Yellow Mottle Virus (RYMV)  
RYMV from West Africa  
RYMV from East Africa  
  
(Trovão *et al.*, 2015) (Dellicour *et al.*, 2018) ⌵ ⌵  
RYMV from Madagascar  
West Nile Virus (WNV)  
WNV from North America

Preferences ⓘ  
☰ ☷  Automatic  
Phylogenetic tre ▾

reloaded  
EvoLaps v.35

TRANSITIONS ⓘ

Transitions without clustering ⓘ  
Display  
Play Slow motion: - 10 +  
Step N: 3  Sliding  Full H  
Brush Off

Transitions with clustering ⓘ

Path options ⓘ  
Layout  
  $\delta$ Size   $\delta$ Offset - +  
ll: - + O: - +  
Full / partial path color  
  
Highlight  
From To  
Misc.  
 Color  HPD  Save map  Save tree

reloaded  
EvoLaps v.35

EDITION ⓘ

Geographic map ⓘ  
Layout  
Positron Labels ▾  
Locations  
Sampled  
Size: - + O: - +  
Ancestral  
Size: - + O: - +  
 HPD Threshold: 80% ▾  
Lat: Lng:  
Heat map ⓘ

Phylogenetic tree ⓘ  
Info  
Layout  
Radial Cartesian  
Ladderize:  up  down  
 $\delta$ ↔: - +  $\delta$ †: - +  
ll: - + Font: - +  
Scale  
 Time scale  
 Most recent sampling

Transition tree ⓘ

Migration distance =  $f(t)$  ⓘ

Inter-cluster exchange charts ⓘ

Ancestral character states ⓘ

# Data toolbox

The screenshot shows the EvoLaps v.35 interface with the following sections:

- DATA** (with an information icon):
  - Import** (with an information icon): A file selection button labeled "Choisir un fichier" and "Aucun fichier choisi", and a "Submit" button.
  - Upload** (with an information icon): A file selection button labeled "Choisir un fichier" and "Aucun fichier choisi", and a "Submit" button.
  - Save/Export** (with an information icon): A "Save" button and a "Save as SVG" dropdown menu.
- Demo** (with an information icon): A library of data sets for demos, including:
  - Rice Yellow Mottle Virus (RYMV) from West Africa
  - RYMV from East Africa (with a thumbnail image of a map and tree)
  - RYMV from Madagascar
  - West Nile Virus (WNV) from North America
- Preferences** (with an information icon): Interface settings including a "Phylogenetic tree" dropdown menu and a checked "Automatic" checkbox.

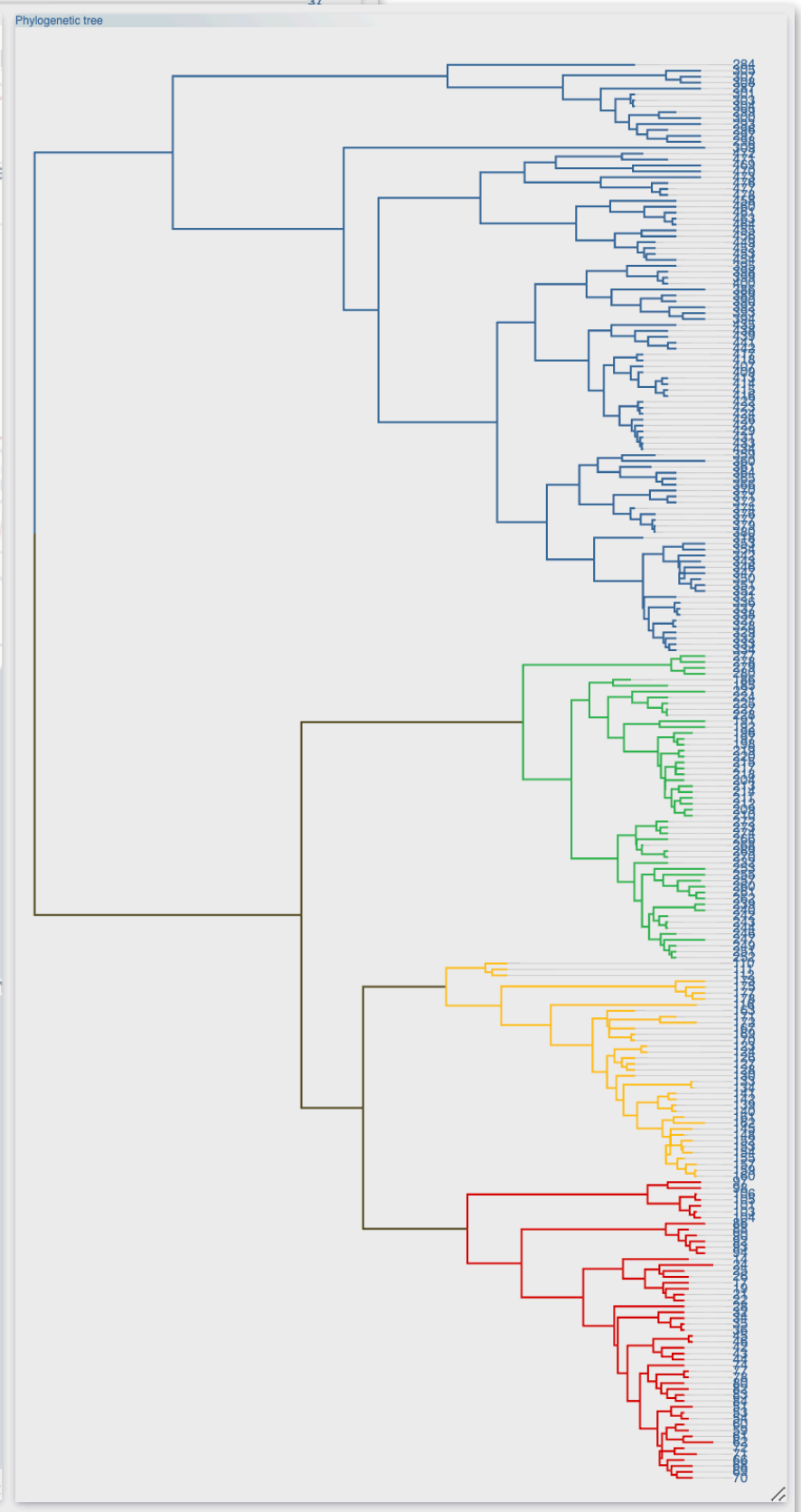
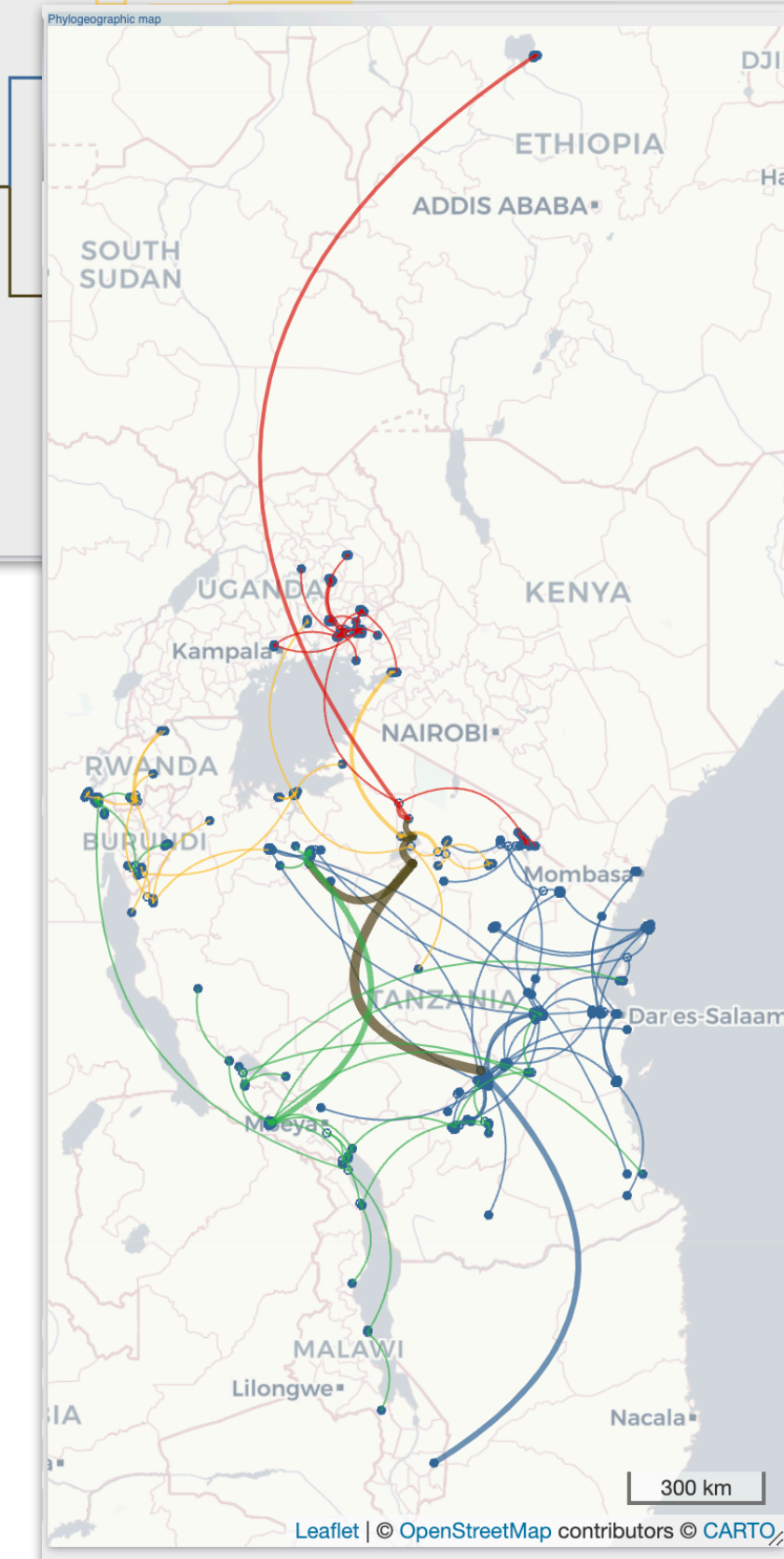
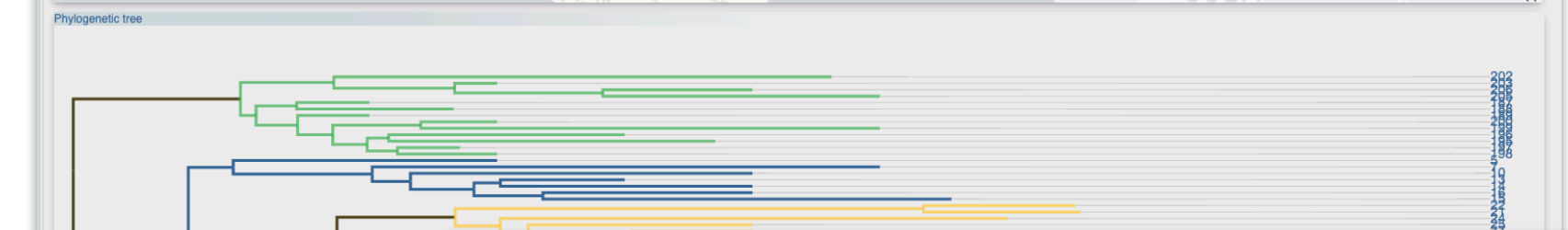
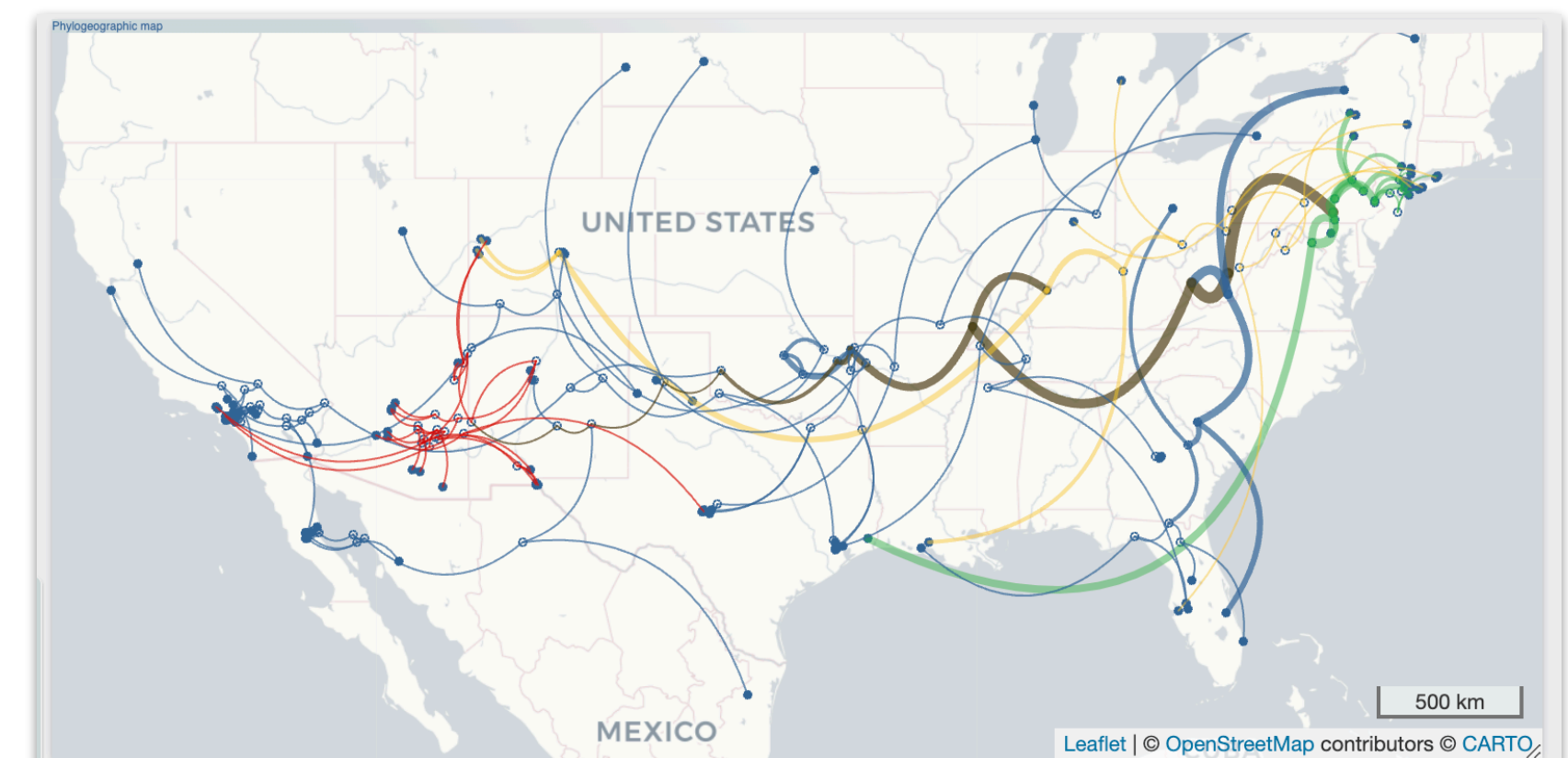
submit a maximum clade credibility (MCC) tree obtained by continuous phylogeographic inference.

upload an evolaps analysis file.

save evolaps analyses or export graphic as SVG files.

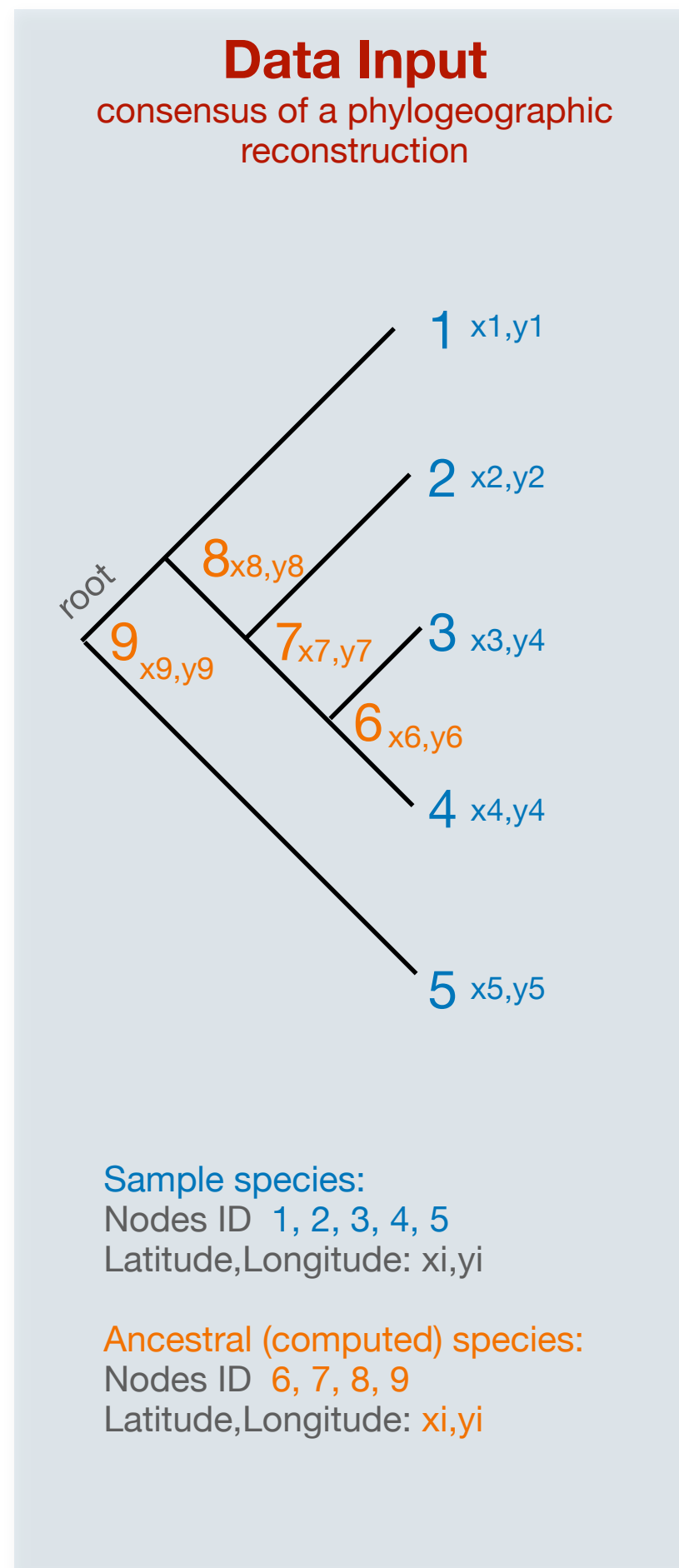
library of data sets for demos.

preferences of the interface. Horizontal/vertical orientation of the geographic map. **Automatic** if checked, each time a graphic component (migration curve, transition tree, transition chart, etc.) is added or removed, the interface is automatically updated to fill the available workspace. The geographic map is oriented from the minima/maxima of latitude/longitude



Horizontal/vertical orientation of the geographic map

# Data toolbox, file formats



submit a maximum clade credibility (MCC) tree obtained by continuous phylogeographic inference. Input data (NEXUS format) must contain a rooted tree with sample and ancestral (consensus) lat/long coordinates for the tips (samples) and the internal nodes of the tree (ancestral species), respectively.

**The variable names saving coordinates must be as following :**

**Samples:**

`location1 location2`  
↑            ↑  
Latitude    Longitude

**Ancestral(\*):**

`location1_median location2_median`  
↑            ↑  
Latitude    Longitude

(\*) if « 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

**HPD:**

`location1_80%HPD location1_90%HPD location1_95%HPD`  
`location2_80%HPD location2_90%HPD location2_95%HPD`

# Transition toolbox, with and without clustering

The phylogeographic scenario can be animated gradually over time, or in time interval steps, over the whole phylogenetic tree or over a particular clade.

The screenshot shows the EvoLaps v.35 interface with the following sections:

- TRANSITIONS** (with an information icon)
- Transitions without clustering**: Includes buttons for Display, Play, Step, and Brush. Settings include Slow motion (10), N: 3, Sliding (checked), Full H (checked), and Brush (Off).
- Transitions with clustering**:
  - Manual clustering**: Radio buttons for Lasso, Clade, and Anchor; Lat/Lng radio button; and M and S sliders.
  - Automatic clustering**:
    - K-means**: K-means button, k: 3.
    - K-means + Lat/Lng**: K-means button, k: 3.
  - Clusters color options**: Clusters section with a clustering-0 slider and a refresh icon.
  - Clusters**: clustering-0: 9 clusters, with Display, Hide, and O: sliders.
- Path options**:
  - Layout**: δSize (checked), δOffset (checked), and II: and O: sliders.
  - Full / partial path color**: Color selection controls.
  - Highlight**: From and To color selection controls.
  - Misc.**: Color (checked), HPD, Save map, and Save tree checkboxes.

- **Display:** the phylogeographic scenario is displayed.
- **Play:** the phylogeographic scenario is animated smoothly at a given speed, thanks to a brush covering the tree. This brush is resizable and moveable.
  - **Slow motion** speed of the animation
- **Step:**
  - **N:** set a number of steps
  - **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 H ON:** the brush height covers the tree height
- **Brush:** see below

# Transition toolbox, brushing the phylogenetic and/or the transition trees

**EvoLaps v.35**

**TRANSITIONS**

**Transitions without clustering**

Display Play Slow motion: - 10 + Step N: 3 Sliding Full H Brush Off

**Transitions with clustering**

Manual clustering Lasso: Clade: Anchor: Lat/Lng: M: S:

Automatic clustering K-means k: 3 K-means + Lat/Lng K-means k: 3

Clusters color options Clusters clustering-0 clustering-0: 9 clusters Display Hide O: +

Display Play Slow motion: - 10 + Step N: 3 Sliding Full H Brush Off

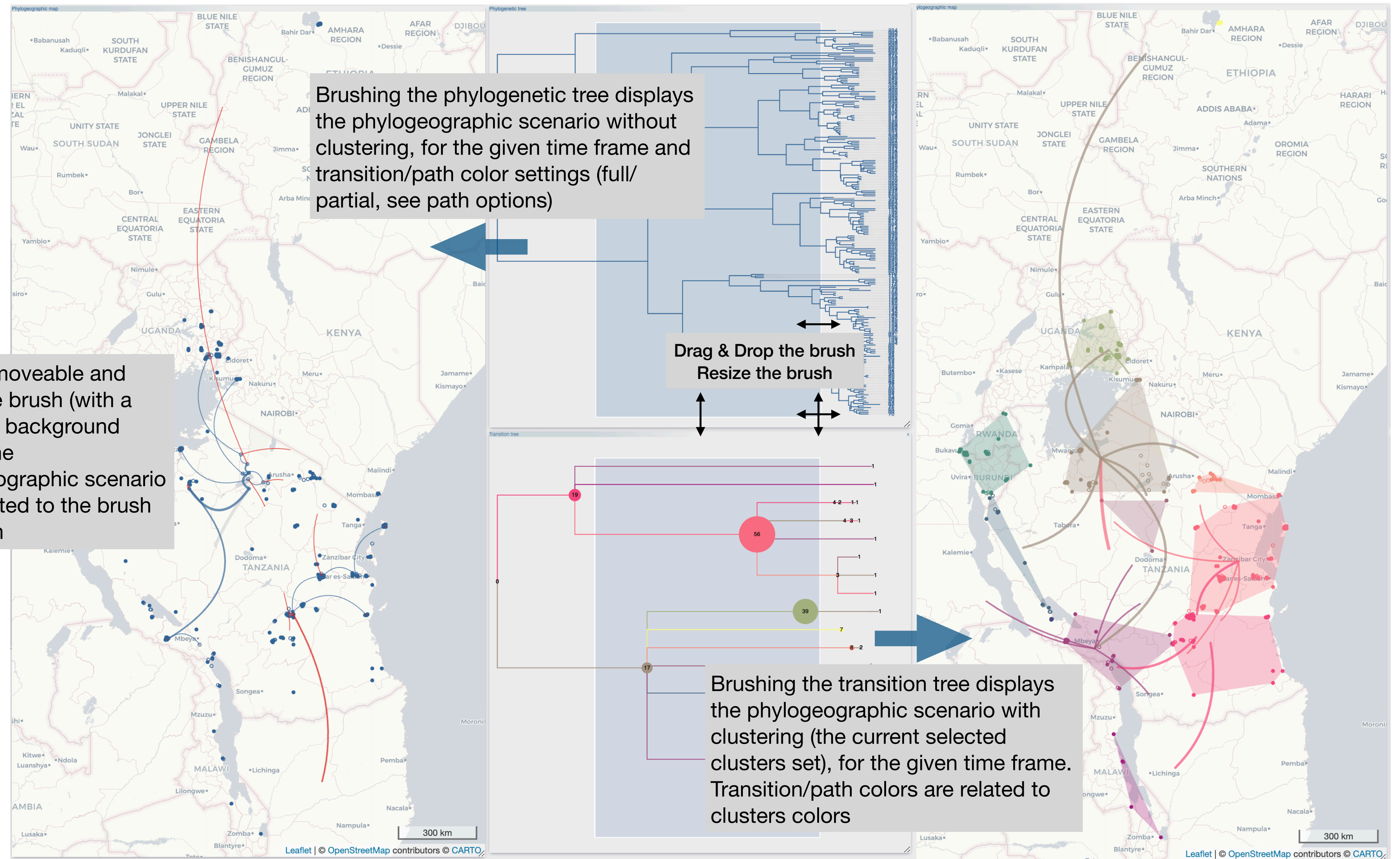
**Path options**

Layout   $\delta$ Size   $\delta$ Offset - + Il: - + O: - +

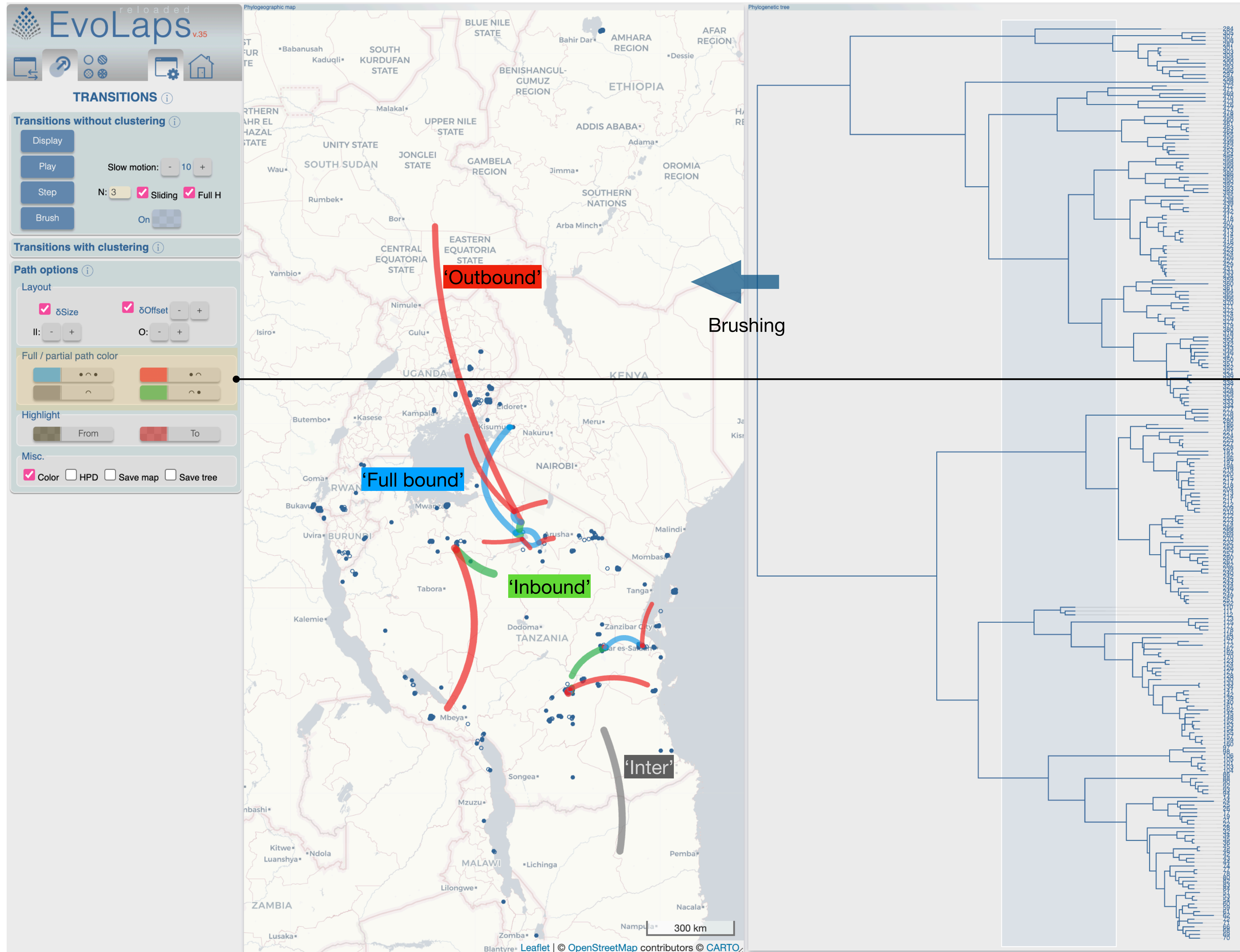
Full / partial path color

Highlight From To

Misc.  Color  HPD  Save map  Save tree



# Transition toolbox, paths options, full/partial paths



path color

**'Full bound':** Color of full bound . If a node and one of its children are inside the brush selection, the corresponding path is « full ».

**'Outbound':** Outbound color. A path is « Outbound » if the corresponding parent node from the tree is inside the brush selection but not the children node

**'Inbound':** Inbound color. A path is « Inbound » if the corresponding children node is inside the brush selection, but not the parent node

**'Inter':** Inter bound color. A path is « Inter » if neither the parent node nor the children are inside the brush selection

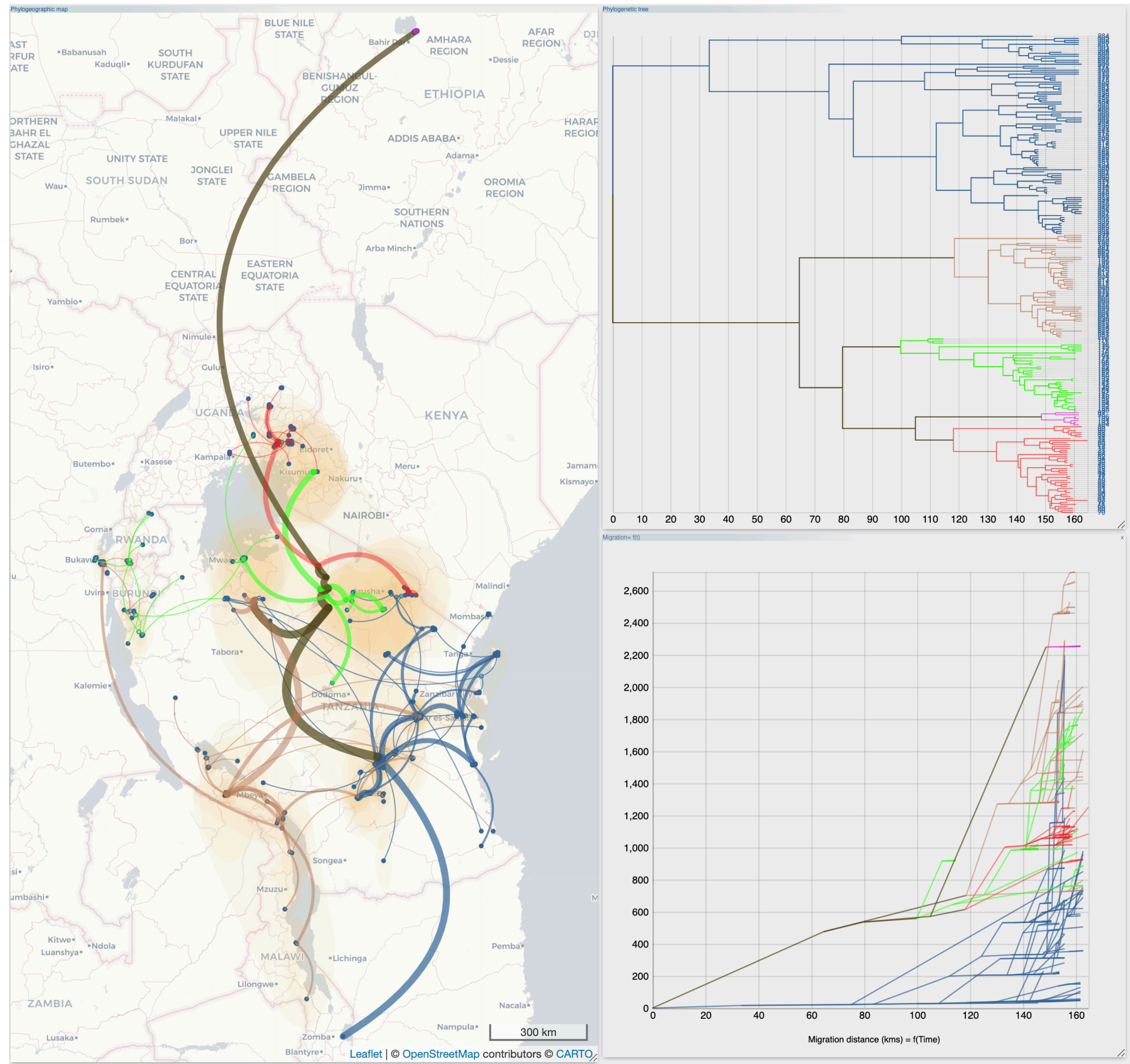


# Transition toolbox, paths options, highlight paths

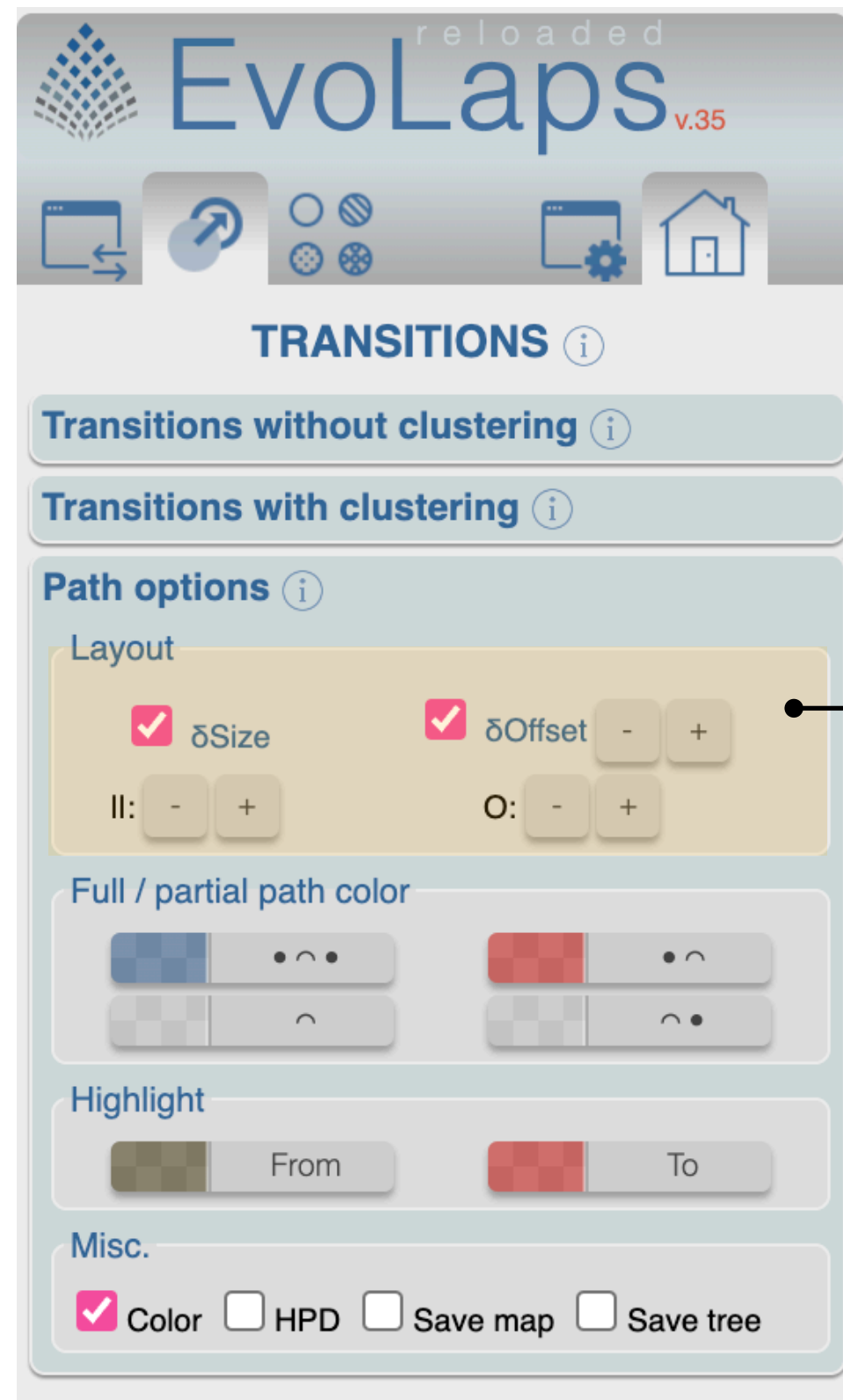
The screenshot shows the EvoLaps v.35 interface with several toolboxes:

- TRANSITIONS** (with an information icon):
  - Transitions without clustering (with an information icon)
  - Transitions with clustering (with an information icon)
- Path options** (with an information icon):
  - Layout**:
    - $\delta$ Size
    - $\delta$ Offset (with minus and plus buttons)
    - Il: (with minus and plus buttons)
    - O: (with minus and plus buttons)
  - Full / partial path color**: (with color selection buttons and arrows)
  - Highlight**: (with 'From' and 'To' color selection buttons)
  - Misc.**:
    - Color
    - HPD
    - Save map
    - Save tree

**Highlight From & To**, highlight color selection (from: backward/to: forward). The coloring of the transitions is done on the fly and in an interconnected way between the graphic components (here the geographical map by pointing paths, the phylogenetic tree by pointing branches and the migration distance curves, by pointing segments).



# Transition toolbox, paths options, layout



Paths options (either with or without clustering)

- $\delta$ Size width path width linked to the depth of the transition in the scenario
- $\delta$ Offset paths curvature linked to the depth of the transition in the scenario,
- $\delta$ Offset -/+ : curvature intensity
- || width of paths. Relative values
- O opacity of paths. Relative values

# Transition toolbox, clustering

The screenshot shows the EvoLaps v.35 software interface. At the top, there's a header with the EvoLaps logo and version number. Below it, a toolbar contains icons for various functions. The main interface is divided into several sections:

- TRANSITIONS**: Contains sub-sections for 'Transitions without clustering' and 'Transitions with clustering'. The 'Transitions with clustering' section is further divided into 'Manual clustering' (with radio buttons for Lasso, Clade, Anchor and sliders for Lat/Lng, M, S) and 'Automatic clustering' (with buttons for K-means and K-means + Lat/Lng, and a 'k' slider).
- Clusters color options**: Includes radio buttons for 'Select' and 'Automatic'. The 'Automatic' section has a 'Set' dropdown (currently 'Steiger') and a color picker. Below it, 'Colors from clusters' shows a small grid of colored squares.
- Clusters**: A list of clusters (clustering-0, clustering-1, clustering-2) with a refresh icon and a trash icon. Below the list are buttons for 'Display', 'Hide', and an 'o' slider for opacity.
- Path options**: Includes buttons for 'Display', 'Play', 'Step', and 'Brush', along with a 'Slow motion' slider and checkboxes for 'Sliding' and 'Full H'.

## Clustering methods

- **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.

## Clusters colors

- Color picker
- 2D color matrices
- 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).

## Clusters list

- submit ans save the current clusters
- cluster selection : display
- reset clustering, display locations without cluster, display/hide clusters, 'o': clusters opacity

# Edition toolbox, geographic map

**EvoLaps v.35**

**EDITION**

**Geographic map**

Layout: Positron Labels

Locations

Sampled: Size: - + O: - +

Ancestral: Size: - + O: - +

HPD Threshold: 80%

Lat: 10.43701530 Lng: 27.37692491

**Heat map**

Choisir un fichier: Aucun fichier choisi

Submit

Gradient: 0.2 0.5 1

Layout: Opa.: - + Blur: - + Rad.: - + Max.: - + Zoo.: - +

Reset Delete

**Phylogenetic tree**

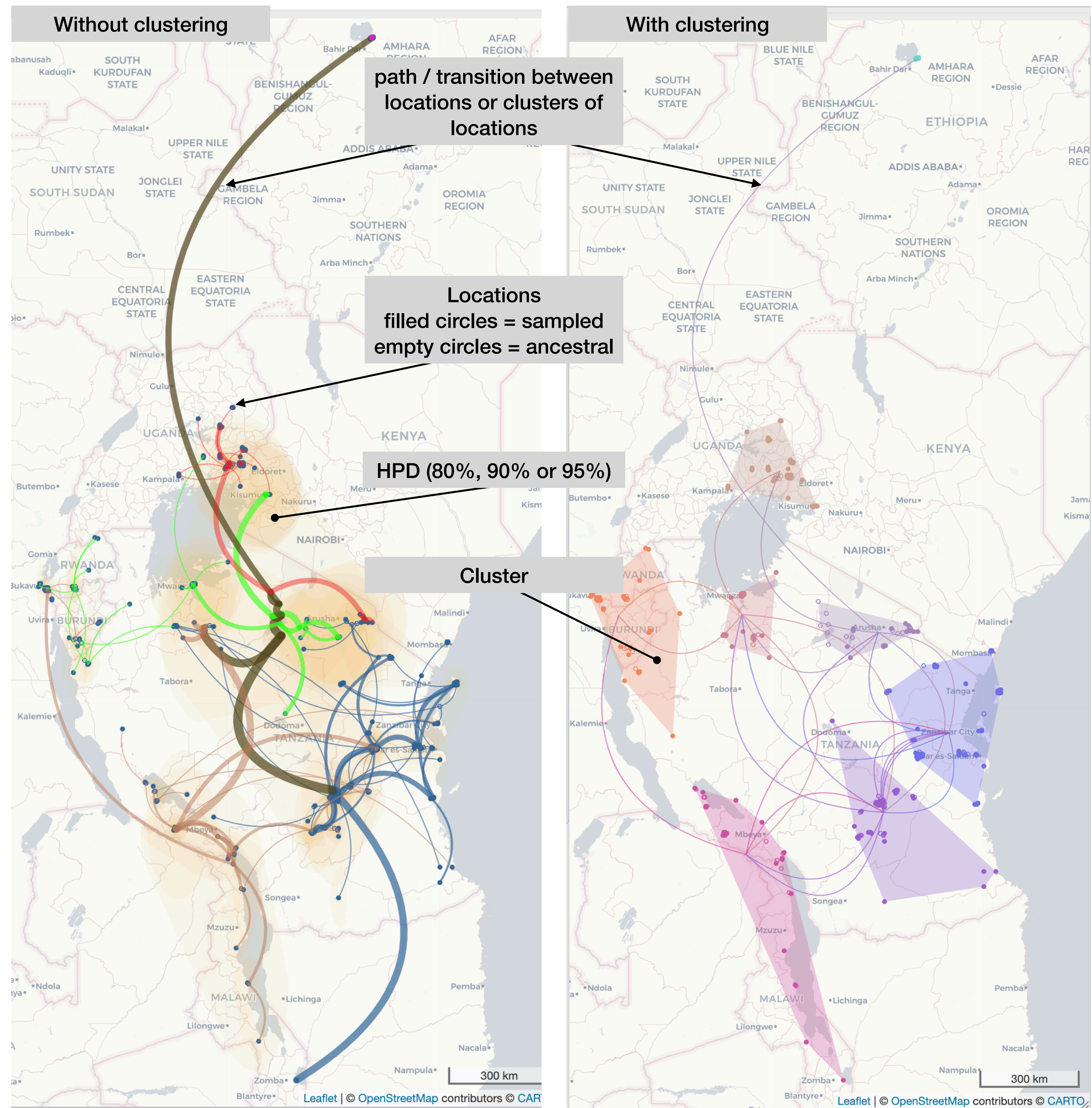
**Transition tree**

**Migration distance = f(t)**

**Inter-cluster exchange charts**

**Ancestral character states**

- **Layout:** map tiles selection
- **Locations:** Size & Opacity of sampled (filled circles) and ancestral (empty circles) locations
- **HPD:** display on/off the Highest Posterior Density, with a selected x% interval (shortest interval in parameter space that contains x% of the posterior probability). Color & Opacity.
- **Lat/Lng:** location of the mouse pointer on the map
- **HeatMap:** see below



# Edition toolbox, geographic map, heat map

**Geographic map**

Layout: Positron Labels

Locations

Sampled

Size: - + O: - +

Ancestral

Size: - + O: - +

HPD Threshold: 80%

Lat: 10.43701530 Lng: 27.37692491

**Heat map**

Choisir un fichier: Aucun fichier choisi

Submit

Gradient

0.2 0.5 1

Layout

Opa.: - + Blur.: - +

Rad.: - + Max.: - +

Zoo.: - +

Reset Delete

Phylogenetic tree

Transition tree

Migration distance =  $f(t)$

Inter-cluster exchange charts

Ancestral character states

CSV file format of the heat map data, the first row must be labelled as following:  
intensity, lat, lng

```
HeatMapData...
intensity, lat, lng
2325,45.361233,11.790552
53,45.31259,9.59238
51,43.14271,11.81649
309,45.90461,9.33011
1063,42.88085,11.67425
270,40.12402,8.823119
53,42.22549,14.01037
2,41.098854,9.524106
636,45.40004,8.9175
320,44.14204,10.66073
4,40.50801,15.81357
197,40.3349,18.29753
12,37.02547,14.49382
37,41.15882,15.33145
2,44.47495,6.99029
100,40.49239,16.15812
12,43.550728,10.507817
```

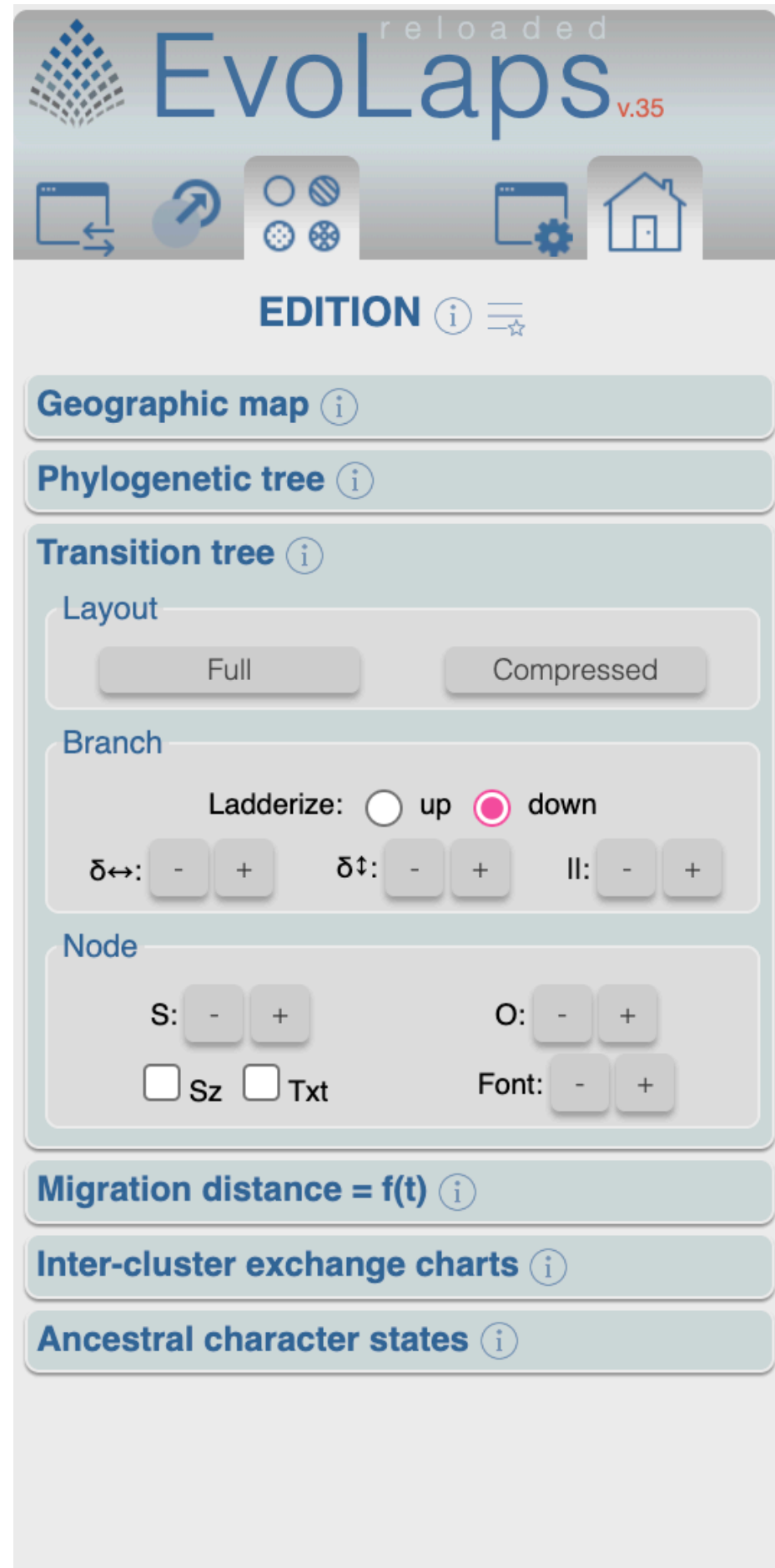
- **Heat map gradient:** threshold and color selection
- **Layout:**
  - **Opa:** minimum opacity the heat will start at
  - **Blur:** amount of blur
  - **Rad:** radius of each 'point' of the heatmap
  - **Max:** maximum point intensity
  - **Zoo:** zoom level where the points reach maximum intensity (intensity scales with zoom)
- **Reset:** reset the gradient and the layout of the heat map
- **Delete:** delete the heat map

# Edition toolbox, phylogenetic tree

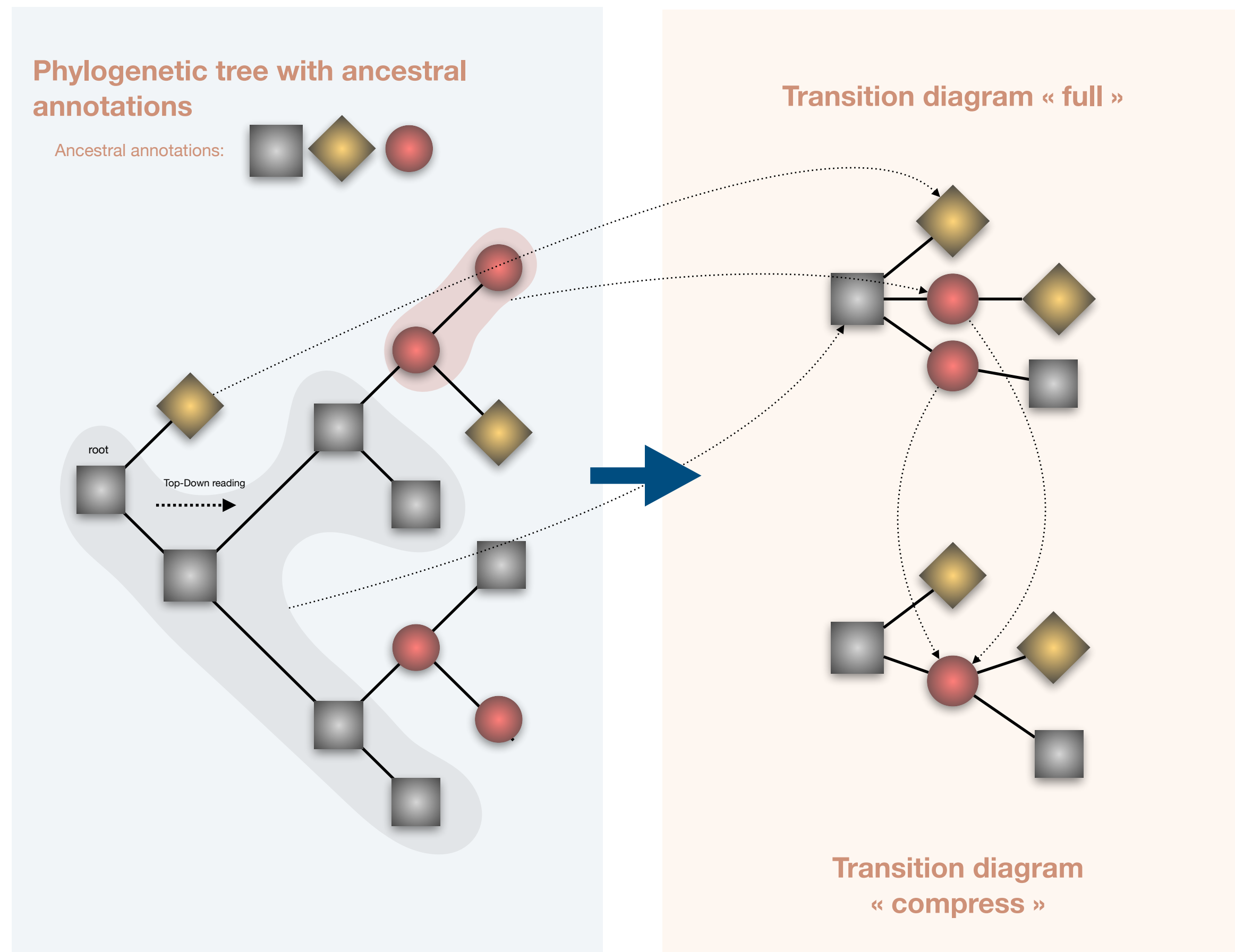
The screenshot shows the 'EvoLaps v.35' software interface. At the top, there is a navigation bar with icons for home, settings, and other functions. Below this is the 'EDITION' menu, which is currently expanded to show several options: 'Geographic map', 'Phylogenetic tree', 'Transition tree', 'Migration distance = f(t)', 'Inter-cluster exchange charts', and 'Ancestral character states'. The 'Phylogenetic tree' option is selected, and its sub-menu is visible. This sub-menu includes an 'Info' section showing '240 leaves, 479 nodes, 478 transitions. Max depth: 18 (160)'. Below the info is a 'Layout' section with buttons for 'Radial' and 'Cartesian', a 'Ladderize' section with radio buttons for 'up' and 'down' (where 'down' is selected), and two sets of minus/plus buttons for adjusting 'δ↔:' and 'δ↑:'. At the bottom of the sub-menu is a 'Scale' section with checkboxes for 'Time scale' and 'Most recent sampling', and a date input field showing 'jj/mm/aaaa'.

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

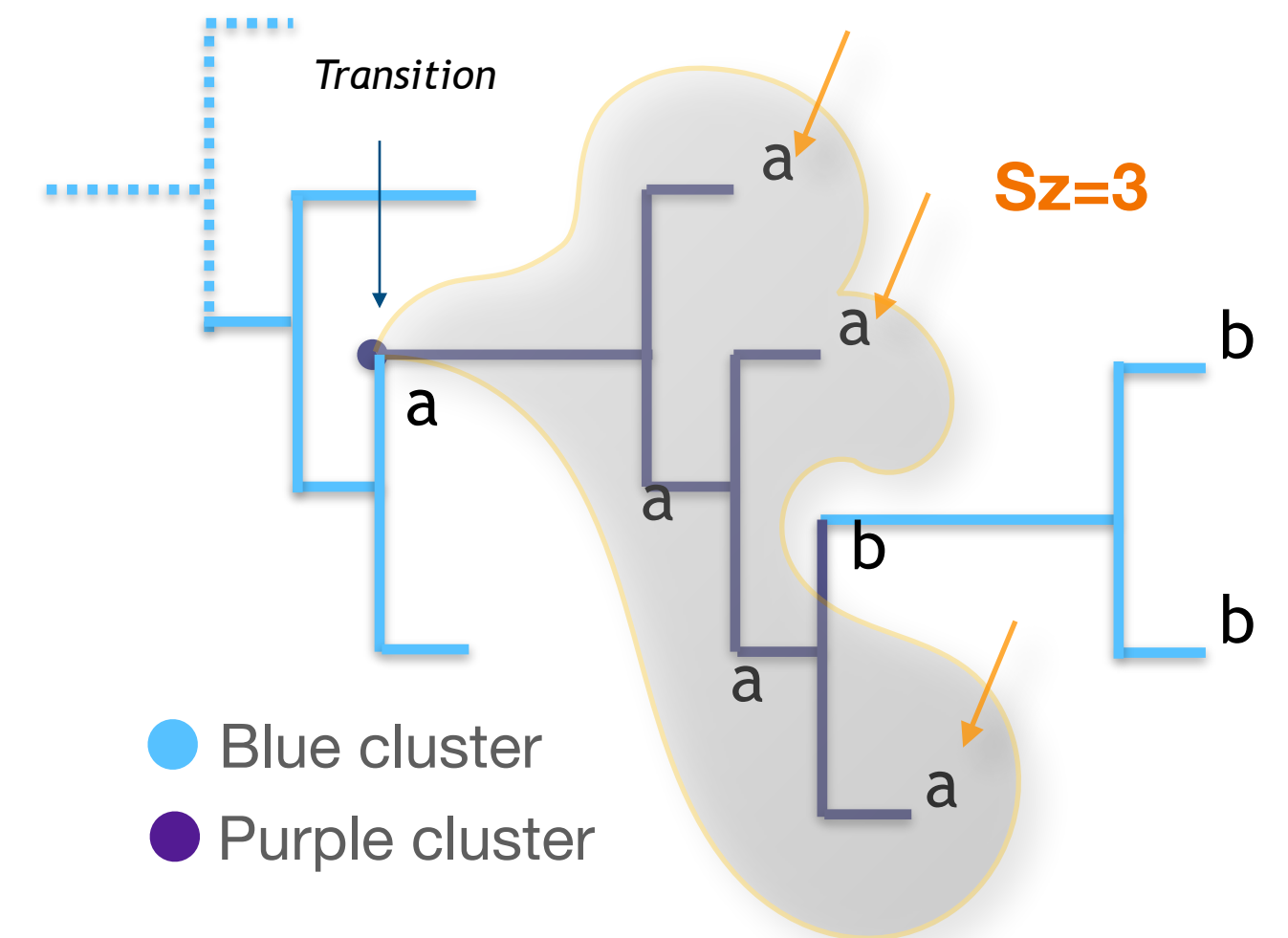
# Edition toolbox, transition tree



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 tree 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 region clusters



Sz 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, Sz values are simply added.

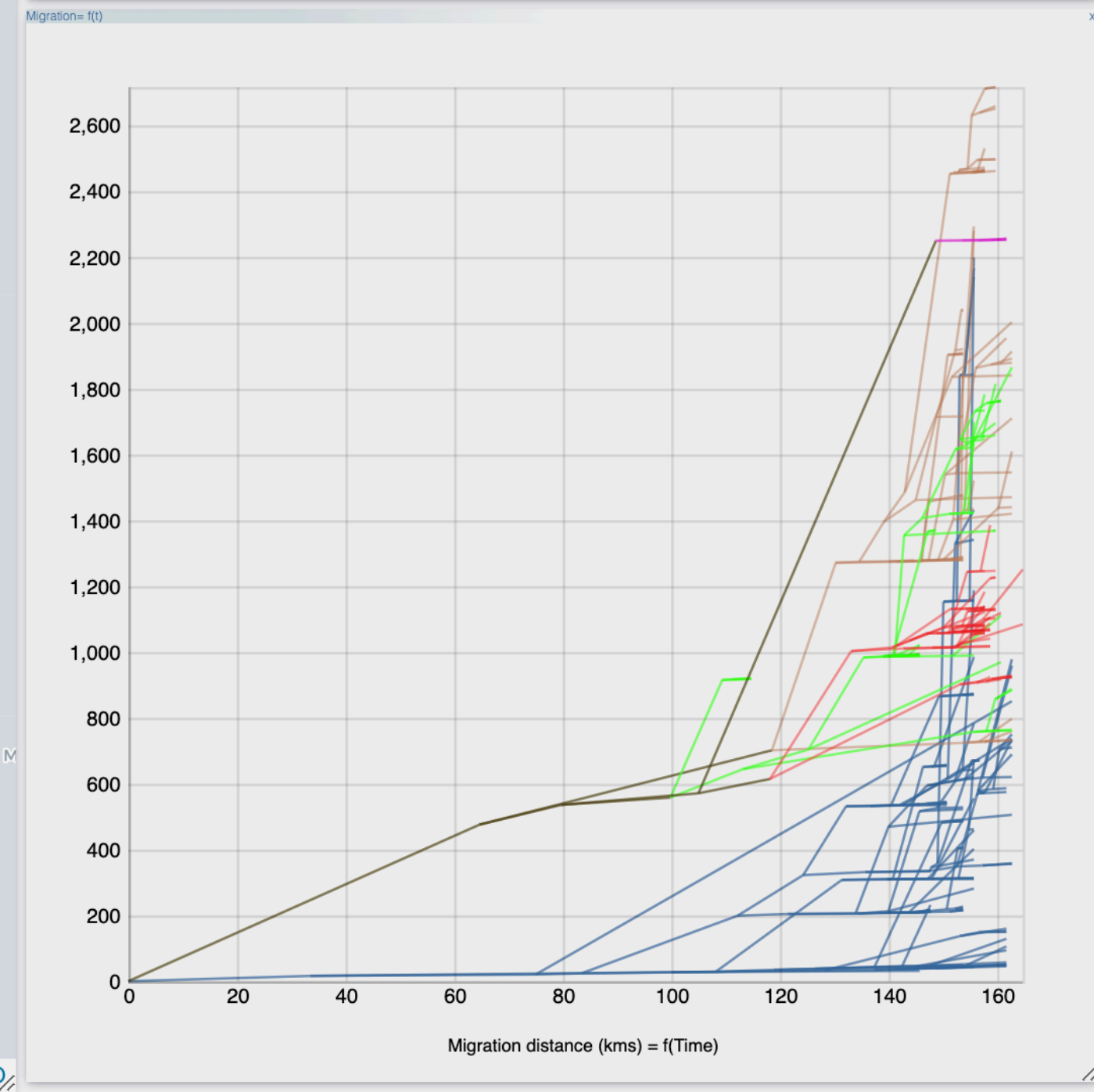
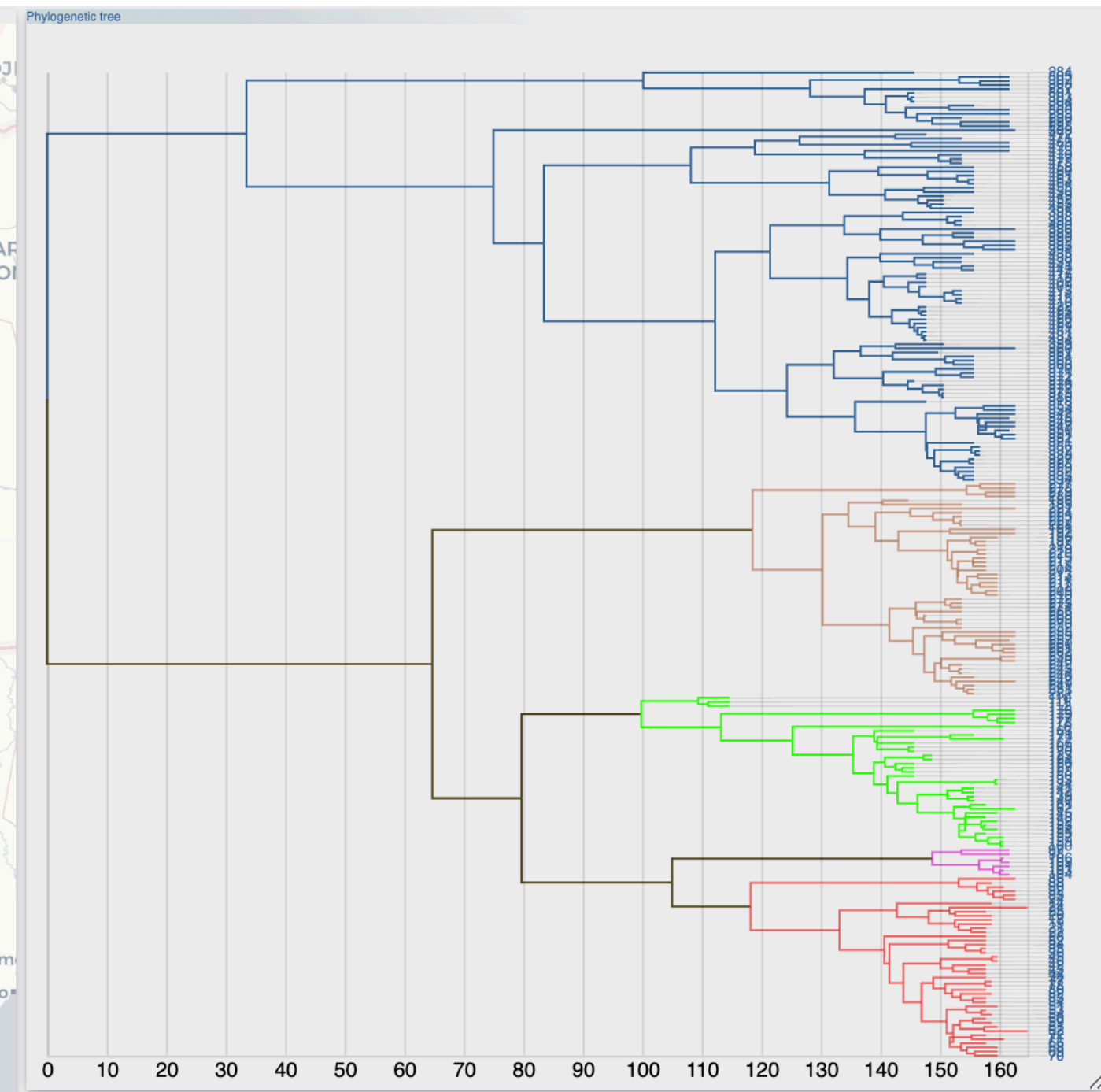
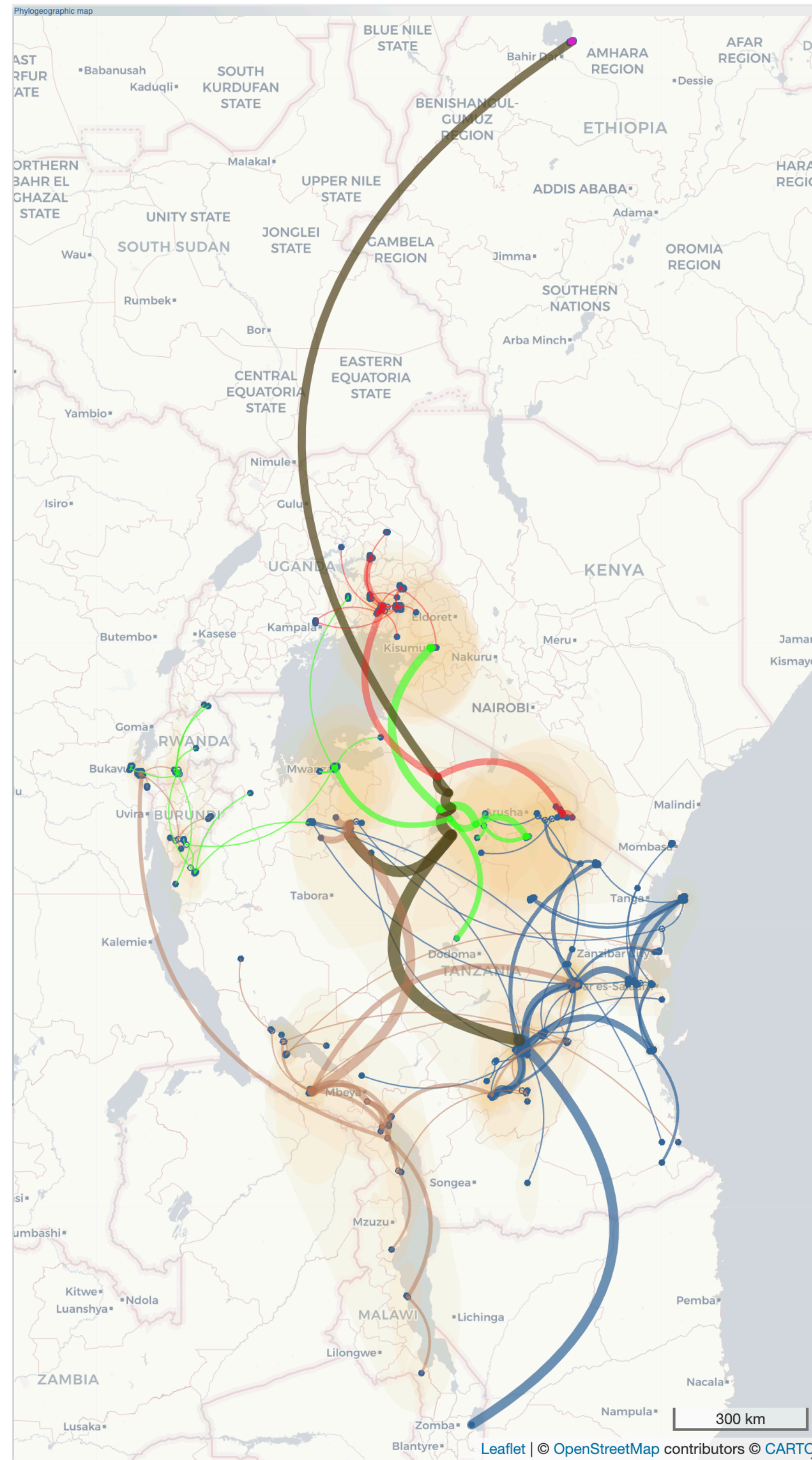


# Edition toolbox, migration distance = f(t)

reloaded  
**EvoLaps** v.35

EDITION ⓘ

- Geographic map ⓘ
- Phylogenetic tree ⓘ
- Transition tree ⓘ
- Migration distance = f(t) ⓘ  
Display
- Inter-cluster exchange charts ⓘ
- Ancestral character states ⓘ



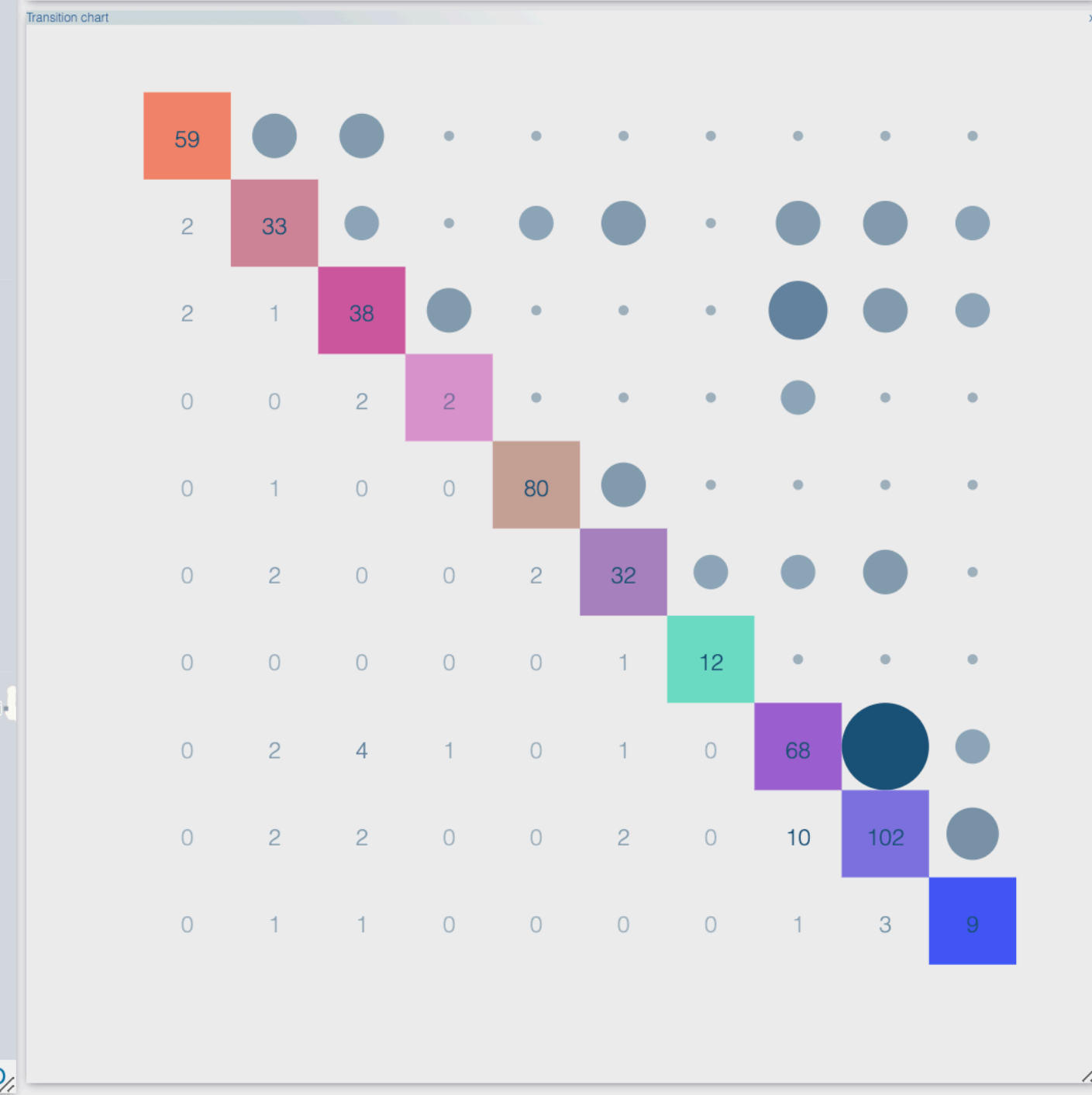
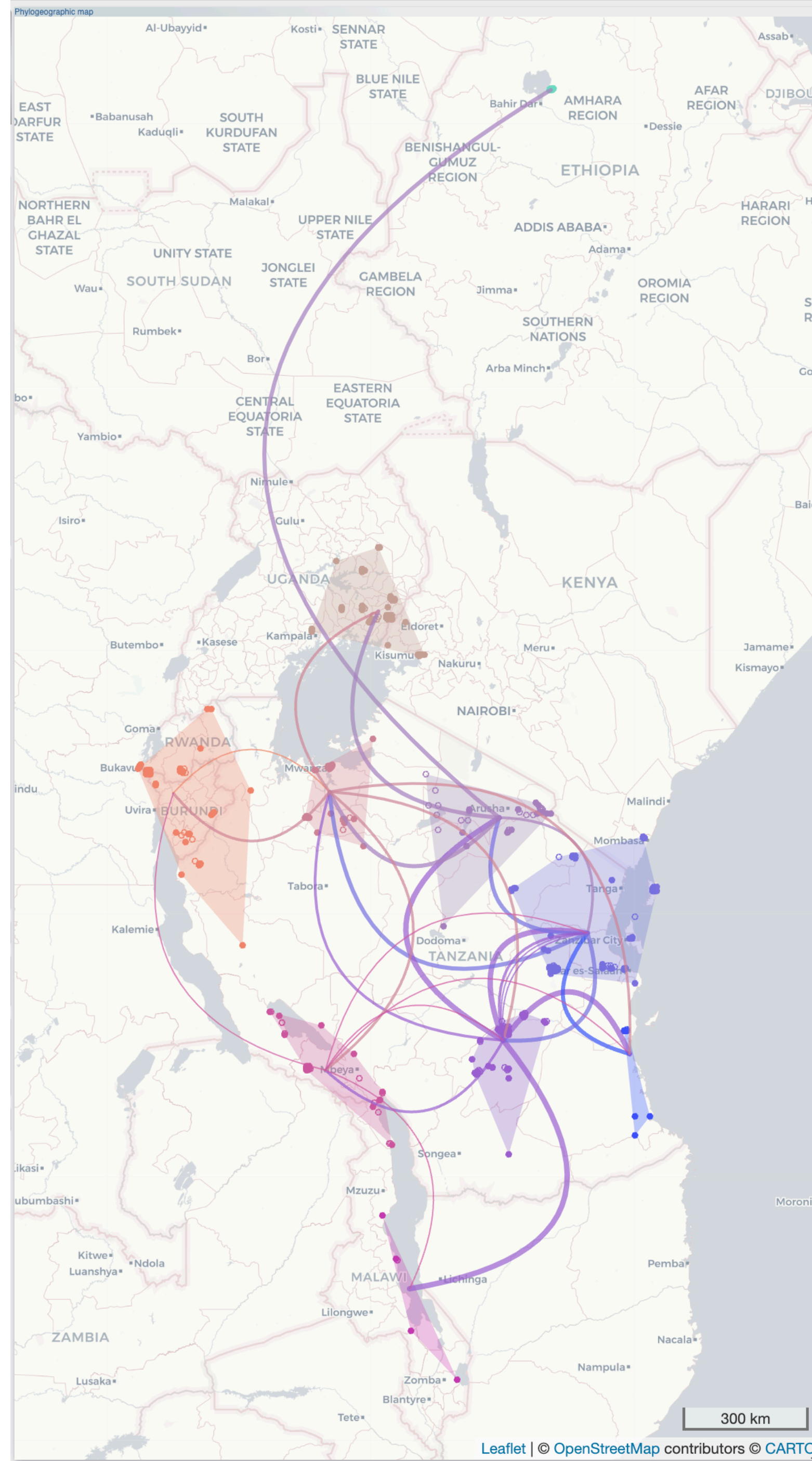


# Edition toolbox, inter-clusters exchange charts

**EvoLaps** reloaded v.35

EDITION ⓘ

- Geographic map ⓘ
- Phylogenetic tree ⓘ
- Transition tree ⓘ
- Migration distance =  $f(t)$  ⓘ
- Inter-cluster exchange charts ⓘ
  - $a \leftrightarrow b$  ⓘ
  - $a \rightarrow b$  ⓘ
  - $a \rightarrow b$  (Lat/Lng) ⓘ
- Ancestral character states ⓘ



# Edition toolbox, ancestral character states

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

reloaded  
EvoLaps v.35

EDITION ⓘ

Geographic map ⓘ

Phylogenetic tree ⓘ

Transition tree ⓘ

Migration distance =  $f(t)$  ⓘ

Inter-cluster exchange charts ⓘ

Ancestral character states ⓘ

Choisir un fichier | Aucun fichier choisi

Model:  F81  Joint

Priors:  Equiprob.  Tree freq.

Submit

Output

Display

Geographic map size - +

Phylogenetic tree size - +

Transition tree size - +

Threshold: M ▼

Color scale: grey single hue ▼

Opacity: - +

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

- 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.

```
DEMO1-EcoType.txt
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
```

## 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  $\pi_i$  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)

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

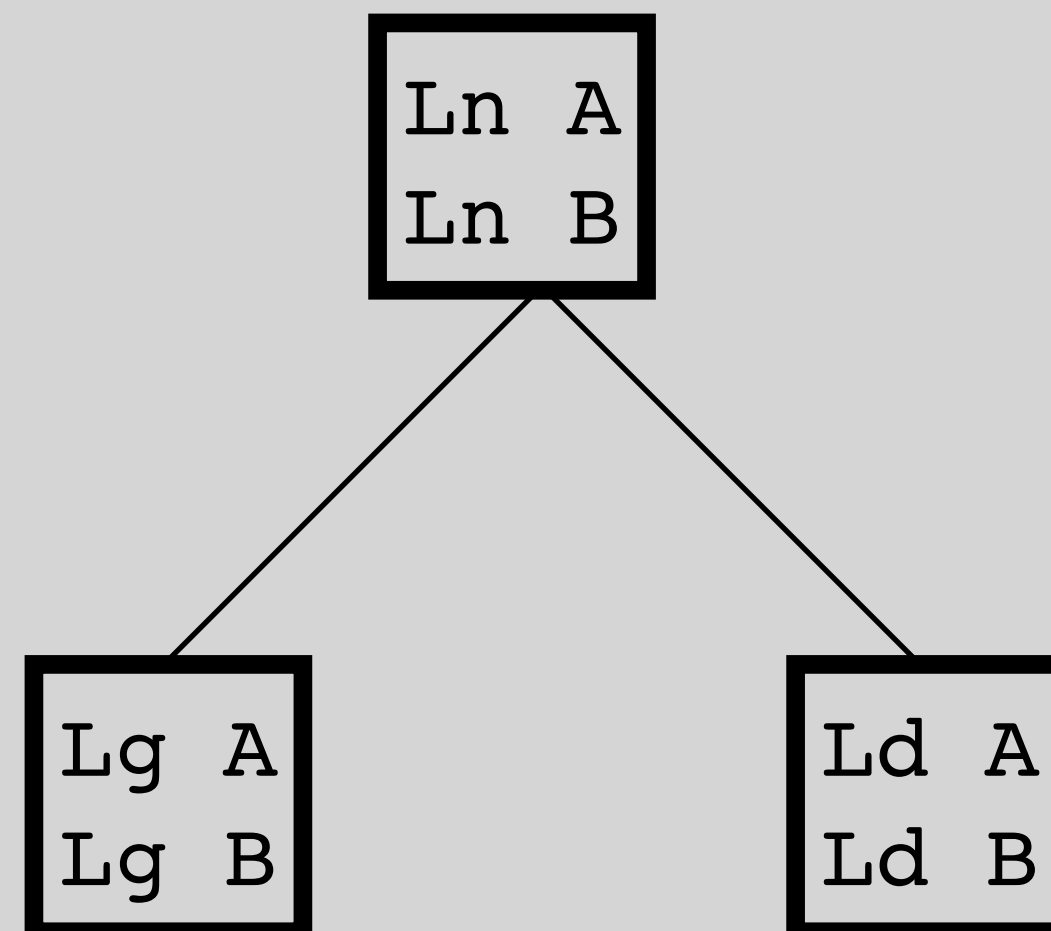
$$L_{\text{down}}(T/i) = \left[ \sum_j PC(i \rightarrow j / l_g) L_{\text{down}}(G/j) \right] * \left[ \sum_j PC(i \rightarrow j / l_d) L_{\text{down}}(D/j) \right]$$

$L_{\text{down}}(F/i) = 1$  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 »

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

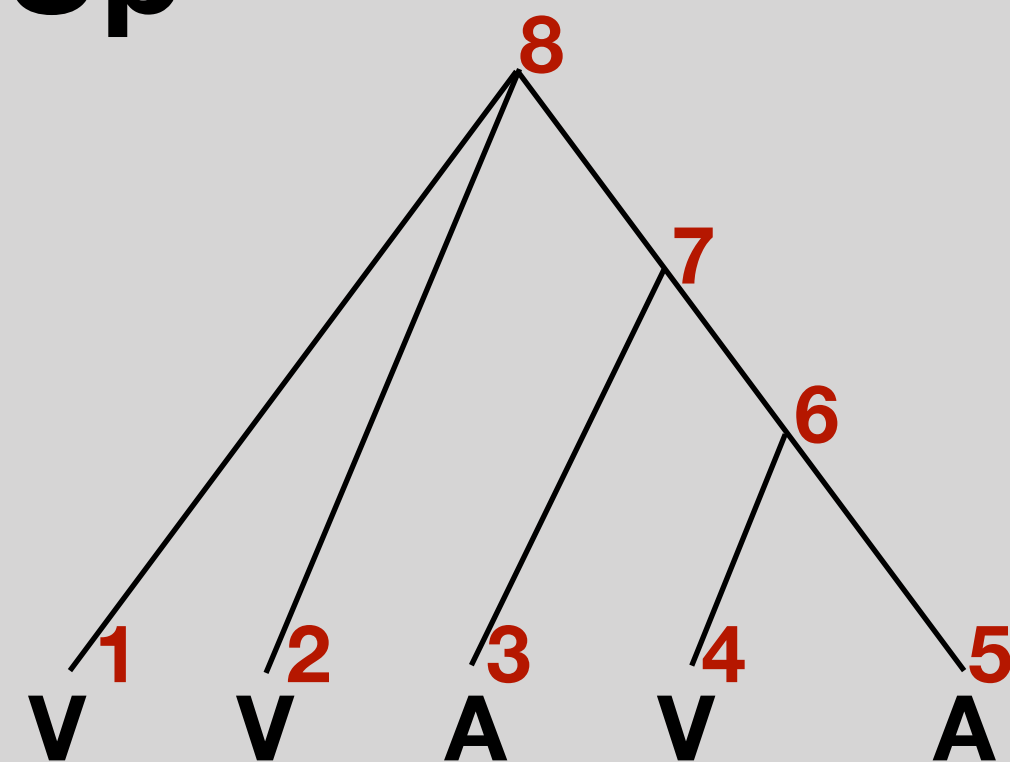
$$L_nB = (L_gA * PC(B \rightarrow A) + L_gB * PC(B \rightarrow B)) * (L_dA * PC(B \rightarrow A) + L_dB * PC(B \rightarrow B))$$

# Computing the joint scenario, example

(Yang, 1995) (Pupko, 2000)

Tree scaling, then : Up + Down

Up



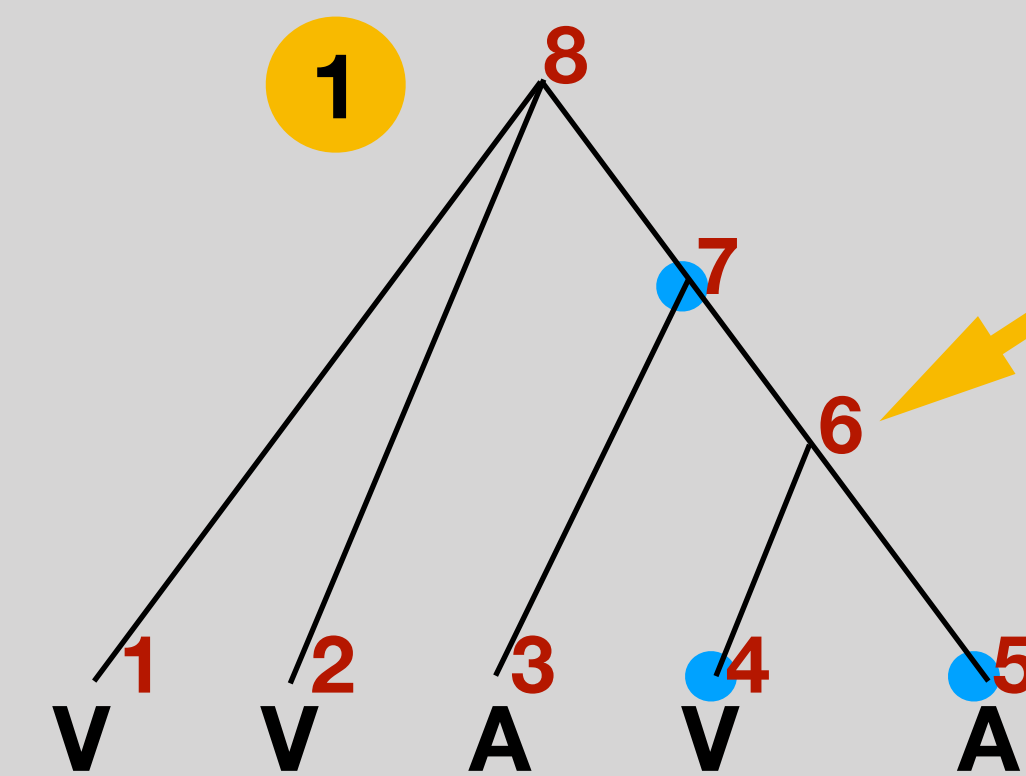
Priors

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

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

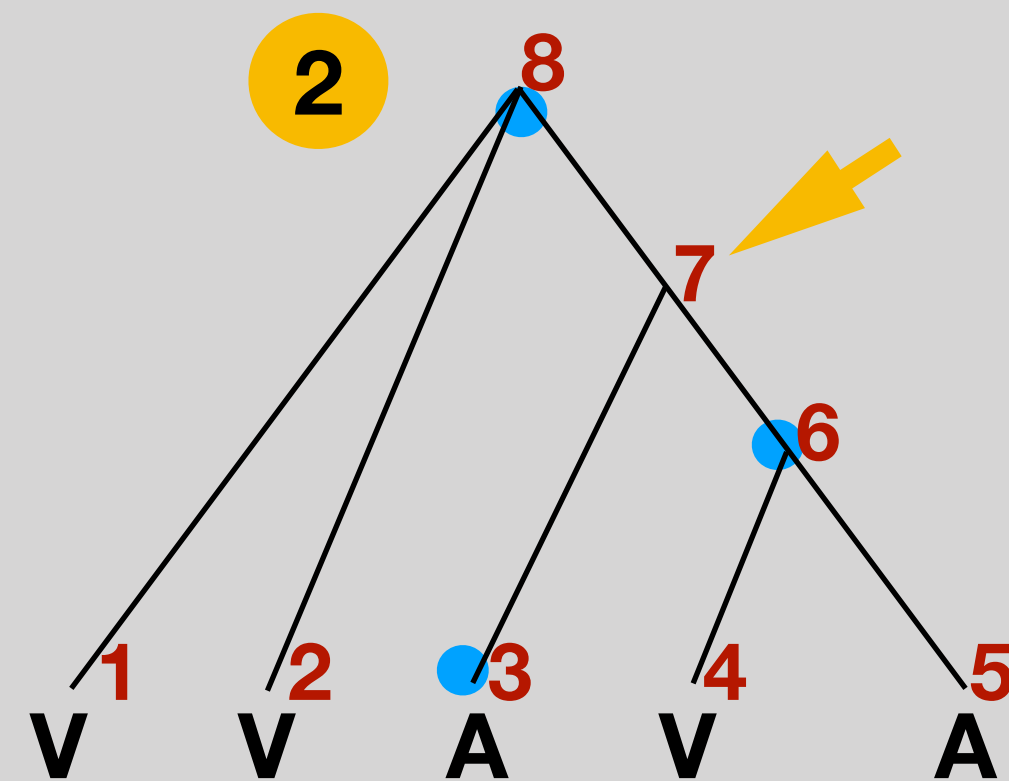
Probabilities of change

PC From	To A	To V
A	0.7	0.3
V	0.45	0.55



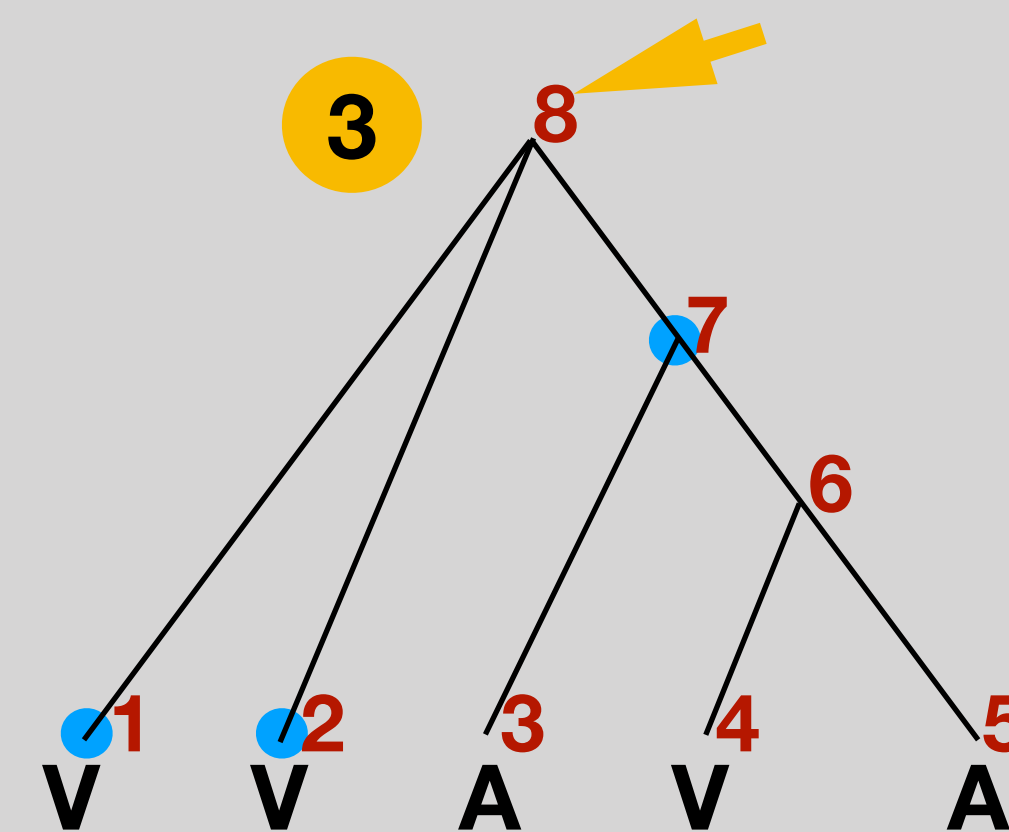
If 7 is A  
 If 6 is A  $L6 = 0.7 \times 0.3 \times 0.7 = 0.147$  → A  
 If 6 is V  $L6 = 0.3 \times 0.55 \times 0.45 = 0.07425$

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  $L7 = 0.55 \times 0.45 \times 0.1361 = 0.03366$



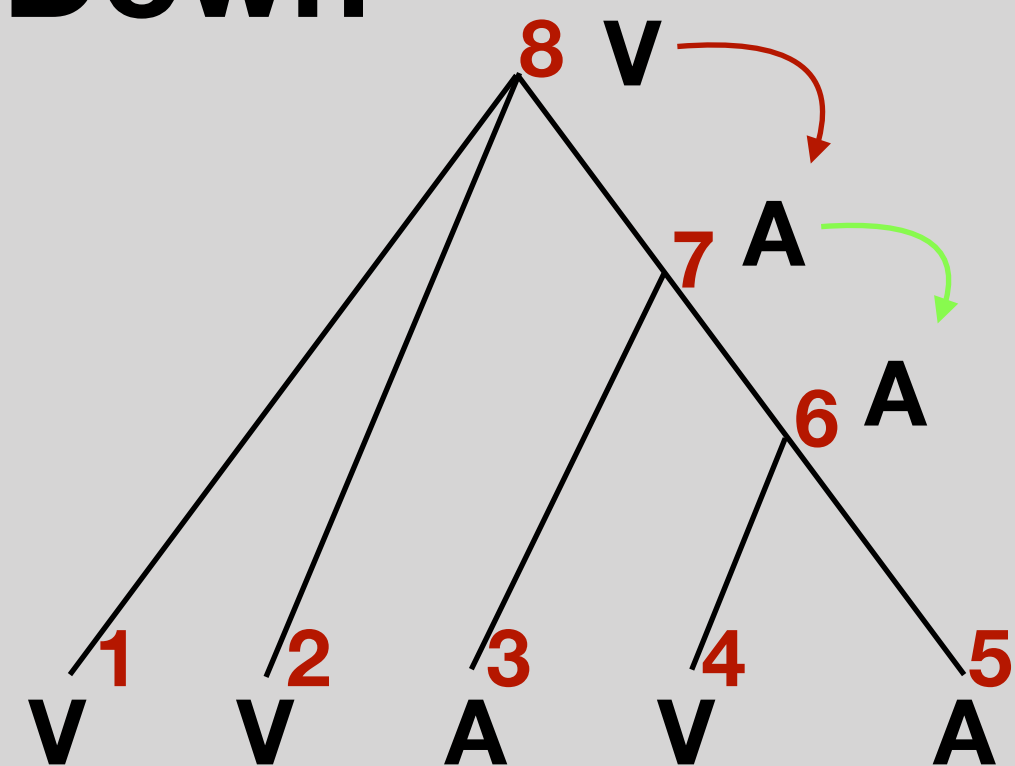
If 8 is A  $L8 = 0.3 \times 0.3 \times 0.072 \times 0.4 = 0.0026$   
 If 8 is V  $L8 = 0.55 \times 0.55 \times 0.0463 \times 0.6 = 0.0084$  → V

# Computing the joint scenario, example

(Yang, 1995) (Pupko, 2000)

Tree scaling, then : Up + Down

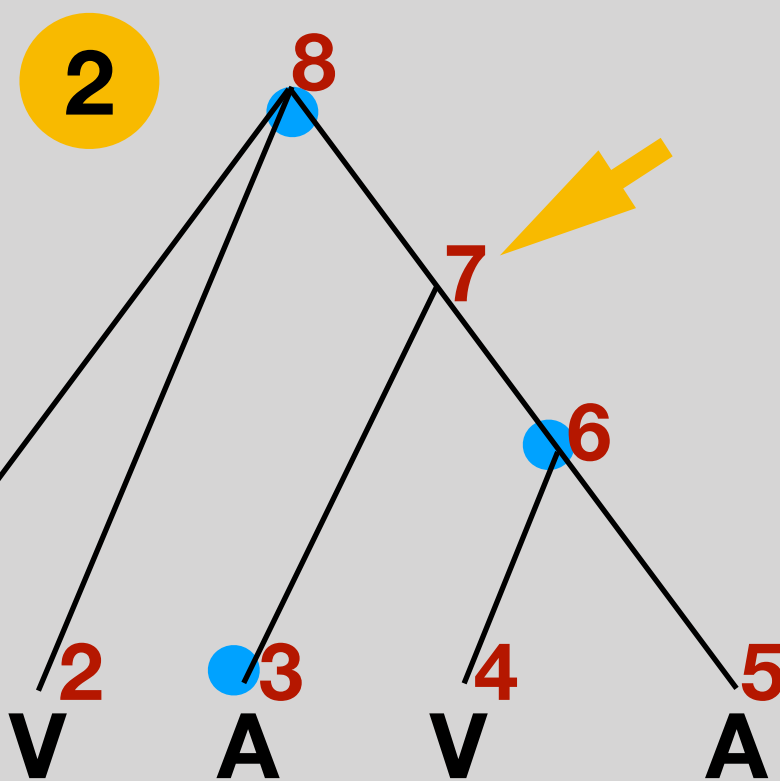
**Down**



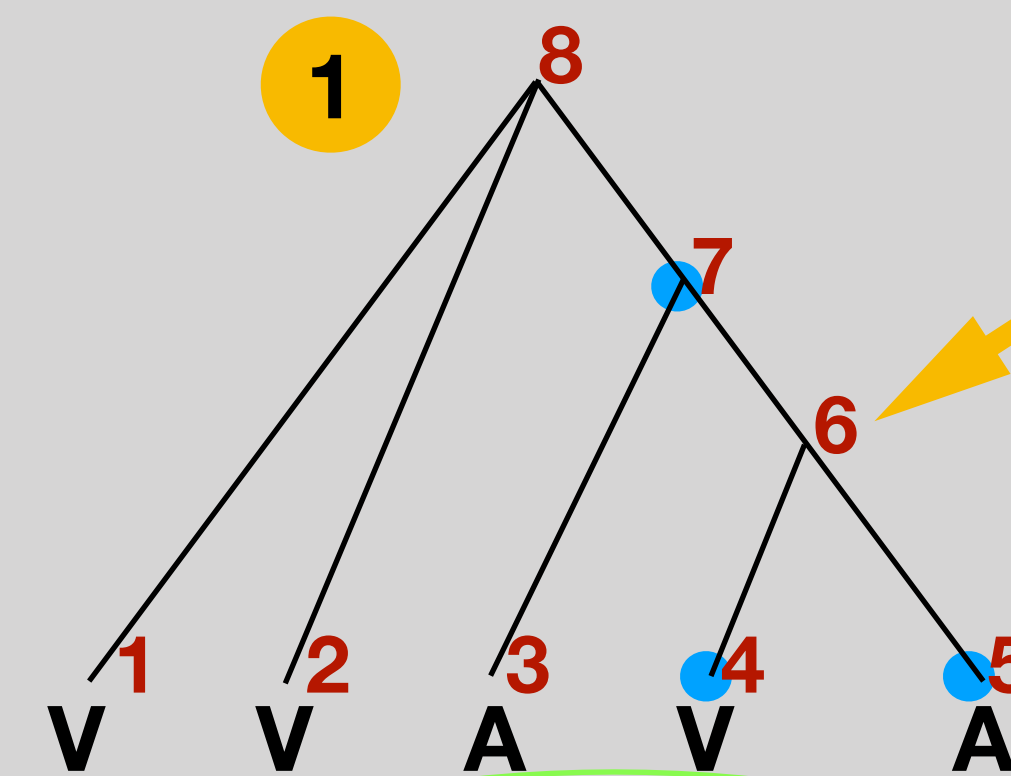
$$\pi_A = 0.4 \text{ (2/5)}$$

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

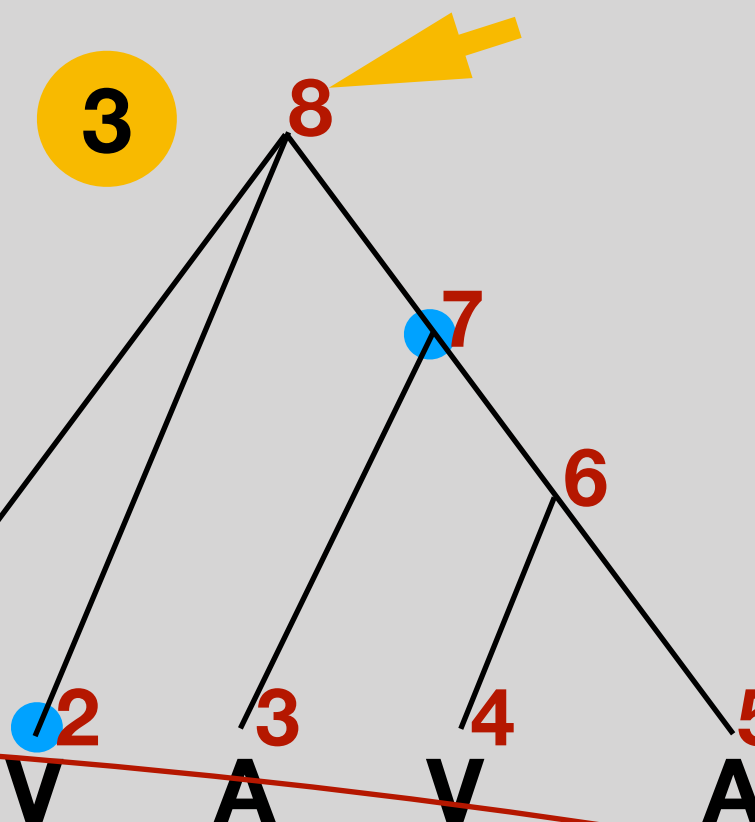
PC	Vers	A	V
De			
A		0.7	0.3
V		0.45	0.55



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