

## 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 from sampled latitude/longitude pairs associated with sampled sequences).

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 clusters of regions using different methods/tools of clustering (lasso, anchor, dynamic latitude/longitude grid, K-means).

The dynamic EvoLaps clustering is iterative, allowing more or less detailed phylogeographic scenarios according to specific spatio-temporal phases of it. 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

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.)EvoLaps...


Rice Yellow Mottle Virus (RYMV)
RYMV from West Africa
RYMV from East Africa

(Trovão et al., 2015) (Dellicour et al., 2018) $\widehat{=}$
RYMV from Madagascar
West Nile Virus (WNV)
WNV from North America

## Preferences

$\square$



## Transition tree

Migration distance $=f(t)$
Inter-cluster exchange charts
Ancestral character states

## Data toolbox



## 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/Ing 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

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

## Sample species:

Nodes ID 1, 2, 3, 4, 5
Latitude,Longitude: xi,yi
Ancestral (computed) species:
Latitude,Longitude: xi,yi
HPD: location1_80\%HPD location1_90\%HPD location1_95\%HPD location2_80\%HPD location2_90\% HPD location2_95\%HPD

## Transition toolbox, with and without clusterins

* EvoLaps

transitions
Transitions 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.

## Transitions with clustering

```
Manual clustering
    Lasso: Clade: © Anchor:
    Latlng: M M Anch
    Latlng: - M:
    Automatic clustering
    K-means
    k-means k:3
    K-means
    K-means + LatLng
    k-means
        Clusters color options
```

    Clusters
    clustering-0 \(\quad \ominus\)
    clustering-0: 9 clusters
    こ. ฉ Display Hide
    \begin{tabular}{|c|c|}
    \hline Display \& <br>
\hline Play \& Slow motion: <br>
\hline \& <br>
\hline
\end{tabular}

        \(\mathrm{N}: 3 \nabla_{\text {Sliding }} \boxtimes_{\text {Full H }}\)
                            Off
    Path options
Path optit
Layout
$\checkmark$ ठSize $\quad \nabla_{\text {бoftset }}$
11: +
Full / partial path color
-•
Highlight
Misc.
$\nabla_{\text {color }} \square_{\text {HPD }} \square_{\text {Save map }} \square_{\text {save tree }}$

- 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:
- $\mathbf{N}$ : 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


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



```
Transition toolbox, paths options, layout
```



TRANSITIONS ${ }_{\text {i }}$
Transitions without clustering
Transitions with clustering
Path options
Layout


Full / partial path color


Highlight
From $\qquad$
Misc.
$\square_{\text {color }} \square_{\text {hpo }} \square_{\text {save map }} \square$ save tree

Paths options (either with or without clustering)

- DSize width path width linked to the depth of the transition in the scenario - $\partial$ Offset paths curvature linked to the depth of the transition in the scenario,

- || width of paths. Relative values
- O opacity of paths. Relative values


## Transition toolbox, clustering

| EvoLaps |
| :---: |
|  |  |

TRANSITIONS

## Transitions without clustering

Transitions with clustering (i)

## Manual clustering



Clusters color options
$\bigcirc^{\text {select }}$


O Automatic


Colors from clusters


Clusters list

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


## 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 paralle//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).
$\mathrm{N}: 3 \nabla_{\text {Sliding }} \nabla_{\text {Full }} \mathrm{H}$ or ${ }^{-1}$

Edition toolbox, geographic map


- 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 $\mathrm{x} \%$ 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



[^0]CSV file format of the heat map data, the first row must be labelled as following:
intensity, lat, Ing
○○○ HeatMapData...
eat 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

```
* EvoLaps.
```


EDITION
Geographic map
Phylogenetic tree i
Info
240 le
240 leaves, 479 nodes, 478 transitions. Max
depth: 18 (160)
Layout
Radial Cartesian
Ladderize: $\bigcirc$ up down
$\delta \leftrightarrow:$
up $\bigcirc$ down
$\delta \uparrow$ :
Font:

Migration distance $=f(t)$
Inter-cluster exchange charts
Ancestral character states

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

Vertical expa

Edition toolbox, transition tree


EDITION ${ }^{(i)}=$


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 $i->j$ 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


Edition toolbox, migration distance $=f(t)$


Edition toolbox, inter-clusters exchange charts

 EDITION (i)

## Geographic map (i)

 Phylogenetic tree Transition tree (i) Migration distance $=\mathrm{f}(\mathrm{t})$ (i) Inter-cluster exchange charts $\underbrace{a<>b}_{a \rightarrow b}$ (i) $\quad$ i $\quad$ a $\rightarrow$ (LatLng) (i) Ancestral character states

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


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 ( $\mathrm{M}=$ 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 >= 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 >= 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 >= 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 1 is expressed as:


```
    = }\mp@subsup{e}{}{-\mul}+(1-\mp@subsup{e}{}{-\mul})\mp@subsup{\pi}{i}{\mathrm{ otherwise}
With \pii equilibrium frequency (priors)
user supplied, roughly estimated from the state frequencies observed
at the tree tips or equal probability
with }\mu=1/(1-\sum\mp@subsup{\pi}{i}{2})\mathrm{ (normalization factor)
```


## Computing the marginal posteriors probabilities is done in four steps:

1 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 posteriors 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
$L(T)=\sum_{i} \pi_{i} L \operatorname{down}(T / i)$
$\operatorname{Ldown}(T / i)=\left[\sum P C\left(i \rightarrow j / I_{g}\right) \operatorname{Ldown}(G / j)\right] *\left[\sum_{j} P C\left(i \rightarrow j / I_{d}\right) \operatorname{LdOWn}(\mathrm{D} / j)\right]$
Ldown $(F / i)=1$ if the tip $F$ is annotated with i, 0 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.

$$
\begin{aligned}
\operatorname{Lup}(G / i) & =\left[\sum_{j} P C\left(i \rightarrow j / I_{g}\right) \operatorname{Lup}(T / j)\left[\sum_{k} P C\left(j \rightarrow k / I_{d}\right) \operatorname{Ldown}(\mathrm{D} / k)\right]\right] \\
\operatorname{Lup}(D / i) & =\left[\sum_{j} P C\left(i \rightarrow j / I_{d}\right) \operatorname{Lup}(T / j)\left[\sum_{k} P C\left(j \rightarrow k / I_{g}\right) \operatorname{Ldown}(\mathrm{G} / k)\right]\right] \\
\operatorname{Lup}(T / j) & =1 \text { si Tis the whole tree }
\end{aligned}
$$

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

```
Post(N,i)=\mp@subsup{\pi}{i}{}L\operatorname{Ldown(N/i) Lup(N/i) / L(T)}
L(T)= \sum \mp@subsup{\pi}{i}{}}\operatorname{Ldown(N/i) Lup(N/i)
```

$$
L(T)=\sum_{i} \pi_{i} \operatorname{Ldown}(T / i)
$$

$\operatorname{Ldown}(T / i)=\left[\sum_{j} P C\left(i \rightarrow j / I_{g}\right) \operatorname{Ldown}(G / j)\right] *\left[\sum_{j} P C\left(i \rightarrow j / I_{d}\right) \operatorname{Ldown}(D / j)\right]$
Ldown $(F / i)=1$ if the tip $F$ is annotated with i, 0 otherwise


Tree scaling, then : Up + Down

Priors

$$
\begin{aligned}
& \pi_{\mathbf{A}}=0.4(2 / 5) \\
& \pi_{\mathbf{V}}=0.6(3 / 5)
\end{aligned}
$$

| PC | ${ }_{0}$ A | V |
| :---: | :---: | :---: |
| A | 0.7 | 0.3 |
|  |  | 0.55 |


If 8 is $A$
If 8 is A
If 7 is $\mathrm{A} L 7=0.7 \times 0.7 \times 0.147=0.072 \quad$ A If 7 is $V \quad L 7=0.3 \times 0.7 \times 0.1361=0.02856$
If 8 is $V$
If 7 is $A L 7=0.45 \times 0.7 \times 0.147=0.0463-\mathbf{A}$ If 7 is $\mathrm{V} \mathrm{L} 7=0.55 \times 0.45 \times 0.1361=0.03366$

Probabilities of change


If 7 is $A$
If is $A L 6=0.7 \times 0.3 \times 0.7=0.147$,
If 6 is $V L 6=0.3 \times 0.55 \times 0.45=0.07425$ If 7 is $V$
If 6 is $\mathrm{A} L 6=0.45 \times 0.3 \times 0.7=0.0945$ If 6 is $V L 6=0.55 \times 0.55 \times 0.45=0.1361-V$


If 8 is $\operatorname{AL8}=\mathbf{A} .3 \times 0.3 \times 0.072 \times 0.4=0.0026$
If 8 is $V$ L8 $=0.55 \times 0.55 \times 0.0463 \times 0.6=0.0084=\mathbf{V}$

Tree scaling, then : Up + Down


If 8 is $A$
If 7 is $A \quad L 7=0.7 \times 0.7 \times 0.147=0.072 \quad \mathbf{A}$ If 7 is $V$
If 8 is $V$ If 7 is $A \quad L 7=0.45 \times 0.7 \times 0.147=0.0463 \sim \mathbf{A}$ If 7 is $V L 7=0.55 \times 0.45 \times 0.1361=0.03366$


If 7 is $A$
If 6 is $A \quad L 6=0.7 \times 0.3 \times 0.7=0.147 \quad \mathbf{A}$ If 6 is $V L 6=0.3 \times 0.55 \times 0.45=0.07425$
If 7 is $V$
If 6 is $A \quad L 6=0.45 \times 0.3 \times 0.7=0.0945$
If 6 is $V L 6=0.55 \times 0.55 \times 0.45=0.1361-V$



[^0]:    Phylogenetic tree
    Transition tree
    Migration distance $=f(t)$
    Inter-cluster exchange charts
    Ancestral character states

