

Tutorial for the R package **seraphim** 1.0

Randomising tree branch positions considering or not an environmental raster influencing the dispersal locations

Simon Dellicour

July 25, 2025

The present tutorial describes how to use the “treesRandomisation()” function of the R package “**seraphim**” [1, 2] to randomise tree branch positions considering or not an environmental raster influencing the lineage dispersal locations. Specifically, we here illustrate how to randomise the tree branch positions inferred by a continuous phylogeographic reconstruction [3] of rabies virus (RABV) lineages having spread in the North American raccoon population [1, 4].

The R package “**seraphim**” is hosted on GitHub (<https://github.com/sdellicour/seraphim>) and the first step is to install it using the “install_github()” function of the “devtools” package:

```
> install.packages("devtools"); library(devtools)
> install_github("sdellicour/seraphim/unix_OS") # for Unix systems
> install_github("sdellicour/seraphim/windows") # for Windows systems
```

Note that the installation of “**seraphim**” requires the preliminary installation of the following R packages: “ape”, “doMC” (only available for Unix systems), “fields”, “gdistance”, “HDInterval”, “ks”, “phytools”, “raster”, “RColorBrewer”, “rgeos”, and “vegan”. Once installed, the package has to be loaded as follows:

```
> library(seraphim)
```

This tutorial requires the following example files also available on the GitHub repository of the package (<https://github.com/sdellicour/seraphim/tree/master/tutorials>): (i) “Extracted.trees”, a folder containing “csv” tree extraction files, and (ii) the environmental raster “Elevation.rast.tif”, an elevation raster encompassing the study area. Extraction files were obtained by extracting the spatio-temporal information embedded in trees

sampled from the post-burn-in posterior distribution of a continuous phylogeographic inference [3] based on the above-mentioned raccoon RABV genomic dataset (see for instance the tutorial “impact on dispersal velocity” for an example of such an extraction step). In these extraction files, each line corresponds to a specific phylogeny branch that can be considered as a vector defined by its start and end location (latitude and longitude), and its start and end dates (in decimal units). Each phylogenetic branch therefore represents a conditionally independent viral lineage dispersal event [5]. This is out of the scope of the present tutorial but a detailed procedure on how to prepare and conduct a continuous phylogeographic reconstruction using the relaxed random walk (RRW) diffusion model [3] implemented in the software package BEAST 1.10 [6] is available on the BEAST community website (https://beast.community/workshop_continuous_diffusion_yfv).

The “treesRandomisation()” function requires the user to specify (i) the directory where the tree extraction tables are saved (“localTreesDirectory”), (ii) the number of tree extraction tables available in this directory (“nberOfExtractionFiles”), (iii) the environmental raster to consider and on which the tree branches randomisation will be conducted (“envVariable”), (iv) the randomisation procedure to use (“randomProcedure”), (v) the boolean variable “repulsion” specifying if the environmental raster should be treated as an environmental variable repulsing (“TRUE”) or attracting (“FALSE”) the dispersal location of the branches (see below for further detail) or if its environmental values should not be considered in the tree branches randomisation process (“NULL”), (vi) the “overwrite” boolean variable specifying if existing simulation outputs should be overwritten or not (if set to “TRUE” and if output files already exists for the simulation $n^{\circ}i$, this simulation will not be conducted and the function will directly conduct simulation $i+1$), and (vii) the “showingPlots” boolean variable specifying if visualisations the different randomisations have to be displayed (and in this case automatically saved) or not.

```
> localTreesDirectory = "Extracted_trees",
> nberOfExtractionFiles = 100
> envVariable = raster("Elevation_rast.tif")
> randomProcedure = 3
> repulsion = NULL
> overwrite = TRUE
> showingPlots = TRUE
```

Several randomisation procedures of tree branches have been implemented: (i) a randomisation of the branch positions while maintaining their length, the tree topology, and the location of the most ancestral node (procedure $n^{\circ}3$; “randomProcedure” set to “3”), (ii) a randomisation of the branch positions while maintaining their length and the location of the starting node of each branch (procedure $n^{\circ}4$; “randomProcedure” set to “4”), (iii) a randomisation of the branch positions while maintaining their length and the location of the ending node of each branch (procedure $n^{\circ}5$; “randomProcedure” set to “5”), and (iv) a randomisation of the branches position while maintaining only their length (procedure $n^{\circ}6$; “randomProcedure” set to “6”). The latter procedure can be useful, e.g., when analysing non phylogenetically informed movement data like capture-mark-recapture movement data. Note that in each case, branch positions are independently

randomised for each tree. Note that randomisation procedures n°1 (using already performed tree randomisations) and n°2 (sing already performed RRW simulations) are not implemented in this function and correspond to procedures solely implemented in the “spreadFactors” function.

If the “repulsion” parameter is set to “TRUE” or “FALSE”, the environmental raster will influence the tree branch randomisation process, higher environmental values tending to repulse (“TRUE”) or attract (“FALSE”) the dispersal locations of lineages. In practice, each branch position will be randomised 100 times and the environmental values at all the potential randomised child (or “ending”) node locations are recorded. Those environmental values V are then used to specify the probability p_j to randomly pick a potential randomised location j given the environmental value v_j at that location: $p_j = (max(V) - v_j)/(max(V) - min(V))$ if “repulsion” is set to “TRUE” (the environmental raster tending to attract the dispersal location of lineages) or $p_j = (v_j - min(v))/(max(V) - min(V))$ if “repulsion” is set to “FALSE” (the environmental raster tending to repulse the dispersal location of lineages). Important note: because only relevant in those cases, these repulsion/attraction randomisation settings are not available for the randomisation procedures n°3 and 4 (if the randomisation procedure n°5 or 6 is selected, the “repulsion” parameter will be set to “NULL” by the fonction).

Note that when the environmental factor is used to define the tendency of the considered environmental factor to repulse or attract lineages, it cannot include any negative values. We will therefore preliminary set the very few negative elevation values within the raster encompassing the study area to “0”:

```
envVariable[(envVariable[]<0)] = 0
```

Once all those parameters are specified, the “treesRandomisation()” function can be called using different randomisation procedures and while considering or not an effect of the environmental raster acting as a factor tending to repulse or attract viral lineages.

1. Tree randomisations without the environmental factor influencing the dispersal location of lineages

1.1. Randomisations of nodes position while maintaining the branches length, the tree topology and the location of the most ancestral node

```
> randomisationDirectory = "Randomisations/Randomisations_1.1"
> repulsion = NULL; randomProcedure = 3
> treesRandomisation(localTreesDirectory, randomisationDirectory,
  nberOfExtractionFiles, envVariable, randomProcedure, repulsion,
  overwrite, showingPlots)
```

1.2. Randomisations of nodes position while maintaining the branches length and the location of the starting node of each branch

```
> randomisationDirectory = "Randomisations/Randomisations_1.2"
> repulsion = NULL; randomProcedure = 4
> treesRandomisation(localTreesDirectory, randomisationDirectory,
  nberOfExtractionFiles, envVariable, randomProcedure, repulsion,
  overwrite, showingPlots)
```

2. Tree randomisations while simulating a negative impact of the environmental variable tending to repulse lineages

2.1. Randomisations of nodes position while maintaining the branches length, the tree topology and the location of the most ancestral node

```
> randomisationDirectory = "Randomisations/Randomisations_2.1"
> repulsion = TRUE; randomProcedure = 3
> treesRandomisation(localTreesDirectory, randomisationDirectory,
  nberOfExtractionFiles, envVariable, randomProcedure, repulsion,
  overwrite, showingPlots)
```

2.2. Randomisations of nodes position while maintaining the branches length and the location of the starting node of each branch

```
> randomisationDirectory = "Randomisations/Randomisations_2.2"
> repulsion = TRUE; randomProcedure = 4
> treesRandomisation(localTreesDirectory, randomisationDirectory,
  nberOfExtractionFiles, envVariable, randomProcedure, repulsion,
  overwrite, showingPlots)
```

Tree randomisations while simulating a negative impact of the environmental variable tending to attract lineages

3.1. Randomisations of nodes position while maintaining the branches length, the tree topology and the location of the most ancestral node

```
> randomisationDirectory = "Randomisations/Randomisations_3.1"
> repulsion = FALSE; randomProcedure = 3
> treesRandomisation(localTreesDirectory, randomisationDirectory,
  nberOfExtractionFiles, envVariable, randomProcedure, repulsion,
  overwrite, showingPlots)
```

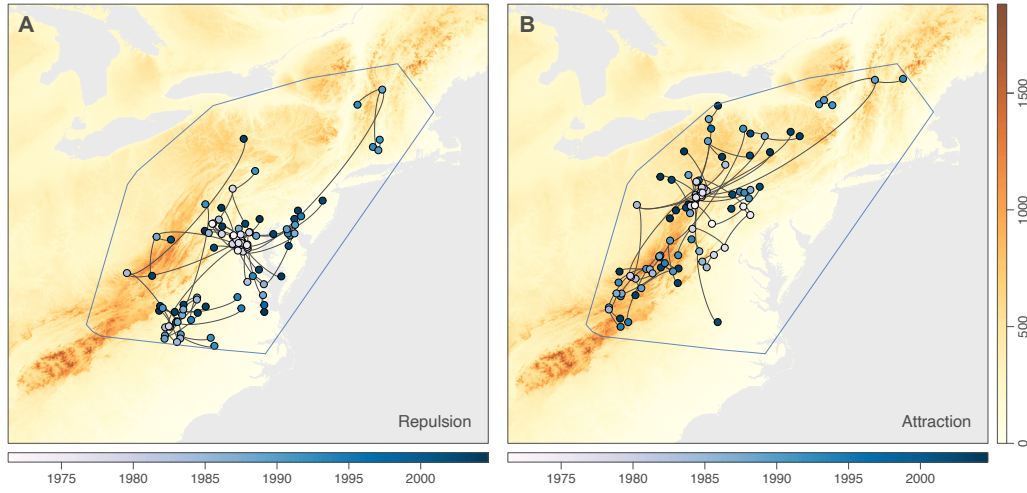


Figure 1: examples of tree branches randomisation obtained using the randomisation procedure n°3 (randomisation of the branch positions while maintaining their length, the tree topology, and the location of the most ancestral node): while simulating a negative impact of the environmental variable tending to repulse lineages (**A**) or a positive impact of the environmental variable tending to attract lineages (**B**). On both maps, the minimum convex hull, which is defined by the node locations of all selected phylogenies and within which the randomisations are conducted, is shown in blue.

3.2. Randomisations of nodes position while maintaining the branches length and the location of the starting node of each branch

```
> randomisationDirectory = "Randomisations/Randomisations_3.2"
> repulsion = FALSE; randomProcedure = 4
> treesRandomisation(localTreesDirectory, randomisationDirectory,
  nberOfExtractionFiles, envVariable, randomProcedure, repulsion,
  overwrite, showingPlots)
```

In Figure 1, we here report the results of one tree branches randomisation obtained using the randomisation procedure n°3 (randomisation of the branch positions while maintaining their length, the tree topology, and the location of the most ancestral node) and while simulating a negative impact of the environmental variable tending to repulse lineages (Fig. ??A) or a positive impact of the environmental variable tending to attract lineages (Fig. ??B).

References

- [1] Dellicour S, Rose R, Pybus OG (2016b). Explaining the geographic spread of emerging epidemics: a framework for comparing viral phylogenies and environmental landscape data. *BMC Bioinformatics* 17: 82.

- [2] Dellicour S, Rose R, Faria N, Lemey P, Pybus OG (2016a). SERAPHIM: studying environmental rasters and phylogenetically-informed movements. *Bioinformatics* 32 (20): 3204-3206.
- [3] Lemey P, Rambaut A, Welch JJ, Suchard MA (2010). Phylogeography takes a relaxed random walk in continuous space and time. *Molecular Biology & Evolution* 27: 1877-1885.
- [4] Biek R, Henderson JC, Waller LA, Rupprecht CE, Real LA (2007). A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. *Proceedings of the National Academy of Sciences of the USA* 104: 7993-7998.
- [5] Pybus OG, Suchard MA, Lemey P, Bernardin FJ, Rambaut A, Crawford FW, Gray RR, Arinaminpathy N, Stramer SL, Busch MP, Delwart EL (2012). Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proceedings of the National Academy of Sciences of the USA* 109: 15066-15071.
- [6] Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution* 4: vey016.