A Rough Guide to RASP 4.2

07/24/2019

Yan Yu[1, *], Christopher Blair[2], AJ Harris[3], Xingjin He[1]

[1] *Key Laboratory of Bio-Resources and Eco-Environment of Ministry of Education, College of Life Sciences,*

Sichuan University, Chengdu, Sichuan, 610065, P. R. China.Email:sculab@gmail.com

[2] *Department of Biological Sciences, New York City College of Technology, The City University of New*

York, 300 Jay Street, Brooklyn, NY 11201, USA.

[3]*Department of Botany, Oklahoma State University, 301 Physical Science, Stillwater, Oklahoma*

74078-3013 USA. Email: aj.harris@okstate.edu

Contents

1. Introduction

1.1 Explanation

We have written RASP (Reconstruct Ancestral State in Phylogenies) software to implements Dispersal-Vicariance Analysis (S-DIVA), Lagrange (DEC model), Bayes-Lagrange, BayArea and Bayesian Binary MCMC (BBM)methods. RASP is easy-to-install on the Windows, Mac, and Linux platforms, provides a user-friendly graphical interface, and generates exportable graphical results.

In RASP, the module of S-DIVA analysis is modified from source code of DIVA 1.2 (Ronquist, 2001) and the MCMC analysis in BBM is modified from source code of Mrbayes 3.1.2 (Ronquist & Huelsenbeck, 2003). DEC model of geographic range evolution was described in Ree et al. (2005) and Ree & Smith (2008). DEC analysis is modified from source code of C++ version of Lagrange developed by Smith (2010). BayArea method was described in Landis et al. (2013)

1.2 Citation

Program:

Yu Y, Harris AJ, Blair C, He XJ. 2015. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical

biogeography. Molecular Phylogenetics and Evolution 87:46-49

S-DIVA method:

Yu Y, Harris AJ, He XJ. 2010. S-DIVA (statistical dispersal-vicariance analysis): a tool for inferring biogeographic histories.

Molecular Phylogenetics and Evolution. 56(2): 848-850

DEC model (Lagrange):

Ree, R H and S A Smith. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction,

and cladogenesis. Systematic Biology. 57(1): 4-14.

S-DEC method

Beaulieu, J.M., Tank, D.C., Donoghue, M.J., 2013. A Southern Hemisphere origin for campanulid angiosperms, with traces

of the break-up of Gondwana. BMC Evolutionary Biology. 13(1), 80.

Bayarea:

Landis M J, Matzke N J, Moore B R, et al. 2013.Bayesian analysis of biogeography when the number of areas is large.

Systematic biology. 62 (6): 789-804

BBM:

Ronquist F, Huelsenbeck JP (2003) MrBayes3: Bayesian phylogenetic inference undermixed models. Bioinformatics 19:1572–1574.

BioGeoBEARS:

Matzke, Nicholas J. (2013). Probabilistic Historical Biogeography: New Models for Founder-Event Speciation, Imperfect Detection, and Fossils Allow Improved Accuracy and Model-Testing. Frontiers of Biogeography. 5(4), 242-248.

Matzke, N.J. (2014). Model Selection in Historical Biogeography Reveals that Founder-event Speciation is a Crucial Process in Island Clades. Systematic Biology. 63(6): 951-970.

Massana, Kathryn A.; Beaulieu, Jeremy M.; Matzke, Nicholas J.; O'Meara, Brian C. (2015). Non-null Effects of the Null Range in Biogeographic Models: Exploring Parameter Estimation in the DEC Model. bioRxiv, http://biorxiv.org/content/early/2015/09/16/026914

Matzke, Nicholas J. (2013). BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. R package, version 0.2.1, published July 27, 2013 at: http://CRAN.R-project.org/package=BioGeoBEARS

R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

BayesTraits:

Pagel M, Meade A. BayesTraits. Computer program and documentation available at http://www.evolution.rdg.ac.uk/BayesTraits.html, 2007.

ChromEvol:

Glick, L., & Mayrose, I. (2014). ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. Molecular Biology and Evolution, 31(7), 1914-1922.

2. Installing [(back to contents)](#)

For Windows users:

Supported Operating Systems: Windows 2000 and above (Recommend Win 10)

System Requirement:

If you are using Windows 2000, 2003 or XP, please make sure that Microsoft® .NET 3.5 Framework is installed on your computer. This is usually installed through Windows updates but may be absent from older systems. The .NET framework 3.5 packages are available for free or you could use the small Client Package for Win XPand should be installed prior to using RASP.

The 4.0 version **NEED** R environment to do many analysis:

If you are beginner and have NO plan to learn anything about R, or you want to use an independent R environment for RASP, you could use the ready-to-use version of RASP 4.0. But the 3rd party package like BioGeoBEARS may not up-to-date.

Usually, we recommend you use the clean version of RASP. You could download R higher than 3.3.3 from its homepage (https://www.r-project.org) or you could simply open [Tools-> Install 3rd Party-> R for Windows]

To install BioGeoBEARS, you could follow the introduction on http://phylo.wikidot.com/biogeobears or simply open [Tools-> Install 3rd Party-> BioGeoBEARS]

If you want to change the version of R using, open [Tools-> Install 3rd Party-> Path of Rscript.exe] to change the path of Rscript.exe.
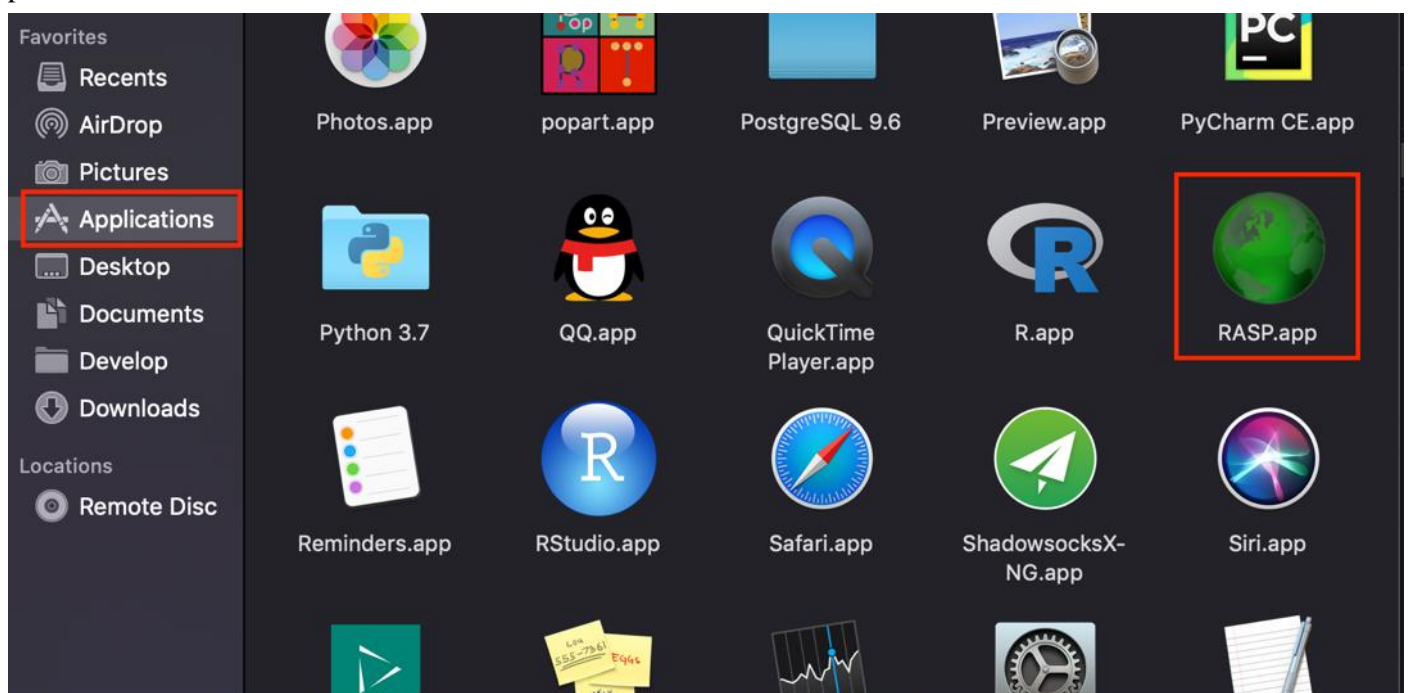
IMPORTANT NOTE FOR ALL PLATFORMS:

Please make sure that the decimal symbol of your system is '.'

For Mac users:

Supported Operating Systems: The last version of macOS

(My MacBook always up-to-date, and I do not test RASP on old version of macOS)

Some user may meet "ERROR! cannot write to Info.plist". Move RASP.app to Applications folder could solve this problem in most of time.

System Requirement:

RASP for Mac uses the Wine program in order for RASP to run.

RASP using the Wineskin wrapper, and Mac OS X's user folders are mapped as following:

Desktop->Desktop

Documents->My Documents

Sometimes, visit "\" may crash RASP for security reason.

Most mac users have reported positive experiences when using RASP. However, we also have received feedback from some Mac users that have experienced stability issues.

If you have problems, you could try to run RASP using virtualization or emulation software. Many options for doing so are available including Parallels (commercial), Crossover (commercial), VM WareFusion (commercial) or VirtualBox (free).

For Linux users:

I think Linux user do not need GUI 😊. You could revise the source code of RASP and compile it on Linux.

If you have problems, you could try to run RASP using virtualization or emulation software.

3. A general tour of RASP

3.1 Preparation

You need:
1. **Trees data set**
   (A sample trees data file: "[RASP folder]\Sample\Rubiaceae\RASP\1000_trees.trees")
   -[How to]: How to obtain trees data set from BEAST (Drummond and Rambaut, 2006)
   -[How to]: How to obtain trees data set from PAUP (Swofford, 2003)
   -[How to]: How to obtain trees data set from other phylogenetic programs.
2. **Condensed tree**
   (A sample condensed tree: "[RASP folder]\Sample\Rubiaceae\RASP\condensed.tre")
   -[How to]: How to make a condensed tree using BEAST.
   -[How to]: How to make a condensed tree using PAUP.
   -[How to]: How to obtain a condensed tree from other phylogenetic programs.
3. **States (Distributions) file (not required).**
   (A sample distributions file: "[RASP folder]\Sample\Rubiaceae\RASP\distribution.csv")
   -[How to]: How to input states (distributions) in RASP.
   -[How to]: How to make a states (distributions) file.

Launch RASP
Note for Mac user: You could put "Sample" folder (and your own data) in Desktop or Documents folder for convenience.
1. Open [**File > Load Trees> Load Trees (more format)**] and navigate to your trees data set and select it.
[Example: Open "[RASP folder]\Sample\Rubiaceae\RASP\1000_trees.trees" in folder Sample]
Note1: In Lagrange, BayArea and BBM analysis, only one condensed tree is needed
Note2: If you have one tree only, you could load it and jump to step 3.
Note3: For a large dataset, one can try [File > Load Trees > Quick Load] to be faster, but only trees exported from Beast and mrbayes were supported.

2. Open [**File>Load Condensed tree>Load User-specified Tree**], navigate to your tree file and select it.
[Example: Open "[RASP folder]\Sample\Rubiaceae\RASP\condensed.tre" in folder Sample]

Do not have a condensed tree: use **[File>Load Condensed tree>Compute Condensed Tree]** to build one.
Note1: The condensed (consensus) tree will be a majority rule consensus with compatible groups with less than 50% support allowed. The tree consensus is computed using Consense (Felsenstein, 1993) from your trees file or from the subset of trees from the file you have specified (i.e., see the Random Trees, and Discard Trees options below).
Note2: The consensus tree is added equal node length by RASP. It could be used for all analysis. However, we strongly recommend user to use their won condensed tree with **branch length**.

3. If you have a states file (distributions file):Open [**File > Load States (Distributions)**], navigate to your file and select it.
[Example: Open "[RASP folder]\Sample\Rubiaceae\RASP\distribution.csv" in folder Sample]
You can also input and revise the states (distributions) in the entry fields in the "State" column in RASP.
If you have multiply states [RASP folder]\Sample\Mutiply states\states.csv]. The actived state will be highlighted as green. You could change the active state by clicking the head of the column.

3.2 Check status

Binary trees: The total number of binary trees in your trees data set.

Amount of trees: The total number of trees in your trees data set.

Note: The **Statistical Dispersal-Vicariance Analysis, Lagrange, Bayes-Lagrange and BioGeoBEARS** requires binary trees while **Bayarea** and **Bayesian Binary MCMC Method** accepts one polytomies tree. **Lagrange**, **Bayes-Lagrange** and **Bayarea** require trees with branch length.

Discard trees: The number of trees that will be discarded from the beginning of the trees data set; equivalent to a burnin.

Random tree: Select random trees form trees dataset to run the **Statistical Method**. Trees will be selected from between **Discard Trees** and **Amount of trees**. You can save the randomly selected trees to a new file with [**File>Export Trees ->Random Trees**] after analysis has been done.

Current Condensed tree: Show the condensed tree in text. You could view your condensed tree in graphic form **[Graphic > Tree View]**

CHECK STATUS: Check the current status of RASP.

Note: We recommend users to click "CHECK STATUS" after any change of their trees or distributions. It will show useful error and warning messages, and tell you which analysis could be done at the current status.

3.3 Reconstruction

We recommend user to **Compare Six Models Using BioGeoBEARS** (see 3.3.1) first, then apply your select method on the **Condensed Tree** (see 3.3.2-3.3.4) or **Trees & Condensed Tree** (see 3.3.5-3.3.7)

3.3.1 Compare Six Models Using BioGeoBEARS

Note: Compare Six Models need R and BioGeoBEARS installed on your computer. You could install them following the manual or you could use ready-to-use version of RASP. We use biogeography of the Hawaiian shrub Psychotria, as a sample dataset. You could load the data from "[RASP folder]\Sample\Psychotria\"

Choose [**Reconstruction > Model Test > Compare Six Models Using BioGeoBEARS**] to show this:
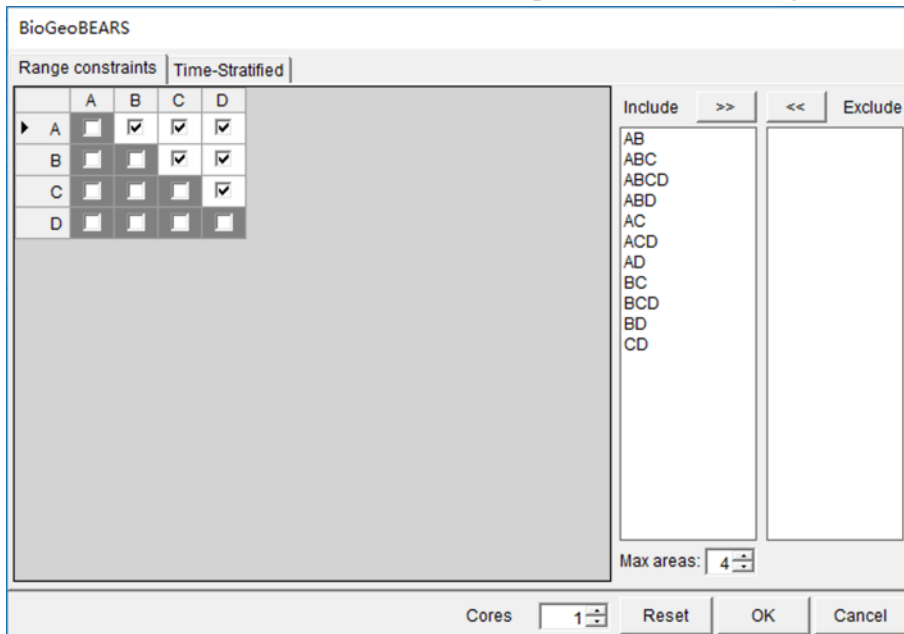
Include -> and Exclude-<: Include or exclude ranges from calculation. This can be accomplished by selecting ranges from the Include list and clicking the arrow to move them to exclude. When you do this, only the selected ranges will be excluded. You can also exclude all ranges that include a particular subset of areas using the range matrix. For example, unchecking the box in the first row, second column will remove all ranges that include A and B; AB and ABC in the example above.

Max areas: The number of unit areas allowed in ancestral distributions.

Cores: You use more cores of CPU to speed up (Only in windows)

Note 1: In BioGeoBEARS, the value of max areas could no less than the maximum number of unit areas of species. For instance, if one species distributed in ABC, then the max areas should >= 3.

Note 2: A command window of R will display after you click OK. Please do **NOT** close it.

The table of the form designates what time slices or periods will be considered. This is important when designating the above matrix if different connectivities/areas/distance are used for different times, and it can also be important when estimating the rate matrix parameters if you want different estimates for different time periods. You could add or remove the periods from the table using the "**Add**" or "**Remove**" button. The left part of the form contains the rate matrix connectivity parameterization.

For instance, the left matrix says that for the period 1 to 2 (In time table, from 0.5 to 1.9) there is less dispersal between areas D and A. And in the period from 0 to 1, all dispersalrates are equal to 1.

You could export/import the matrix to/from file.

[Example: Open "[RASP folder]\Sample\Psychotria\timeperiods.txt"]

3.3.2 Dispersal-Extinction-Cladogenesis (DEC)

Note: We use biogeography of the Hawaiian shrub Psychotria, as a sample dataset. You could load the data from "[RASP folder]\Sample\Psychotria\"

Choose [**Reconstruction >On Condensed Tree >Dispersal-Extinction-Cladogenesis (DEC)**] to use DEC models of geographic range evolution described in Ree et al. (2005) and Ree and Smith (2008). This module use source code of c++ version of Lagrange developed by Smith (2010) and is much more faster than the Python version of Lagrange.

Include -> and Exclude-<: Include or exclude ranges from calculation. This can be accomplished by selecting ranges from the Include list and clicking the arrow to move them to exclude. When you do this, only the selected ranges will be excluded. You can also exclude all ranges that include a particular subset of areas using the range matrix. For example, unchecking the box in the first row, second column will remove all ranges that include A and B; AB and ABC in the example above.

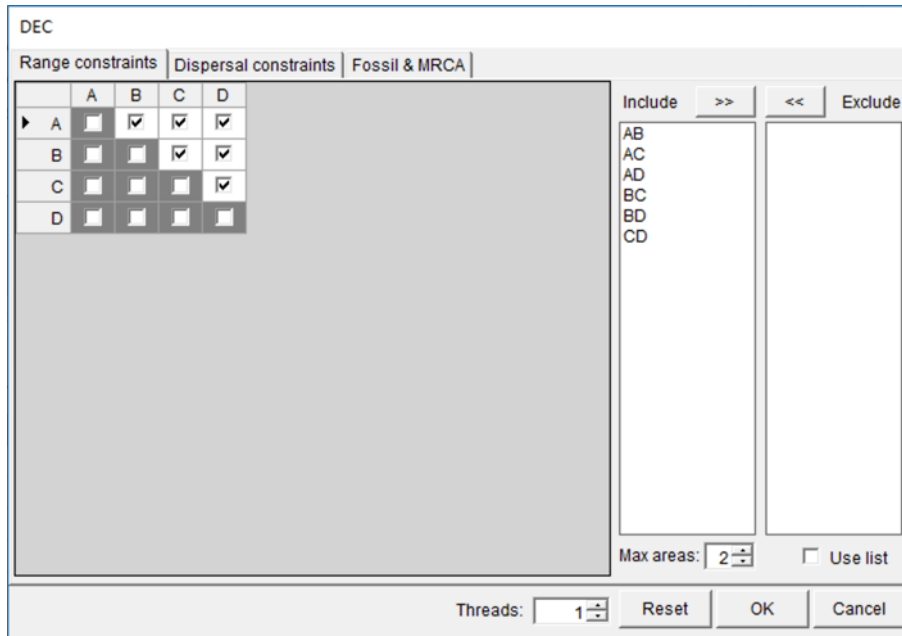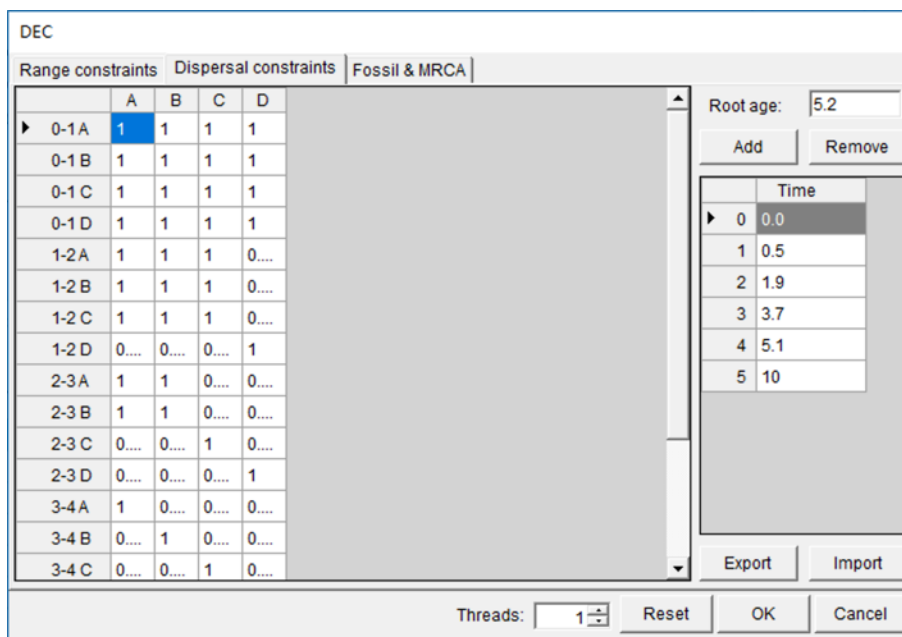**Maximum areas**: The number of unit areas allowed in ancestral distributions.

The right table of the form designates what time slices or periods will be considered. This is important when designating the above matrix if different connectivities are used for different times, and it can also be important when estimating the rate matrix parameters if you want different estimates for different time periods. You could add or remove the periods

from the table using the "**Add**" or "**Remove**" button. The left part of the form contains the rate matrix connectivity parameterization.

For instance, the left matrix says that for the period 1 to 2 (In time table, from 0.5 to 1.9) there is less dispersal between areas D and A. And in the period from 0 to 1, all dispersalrates are equal to 1.

You could export/import the matrix to/from file.

[Example: Open "[RASP folder]\Sample\Psychotria\timeperiods.txt"]
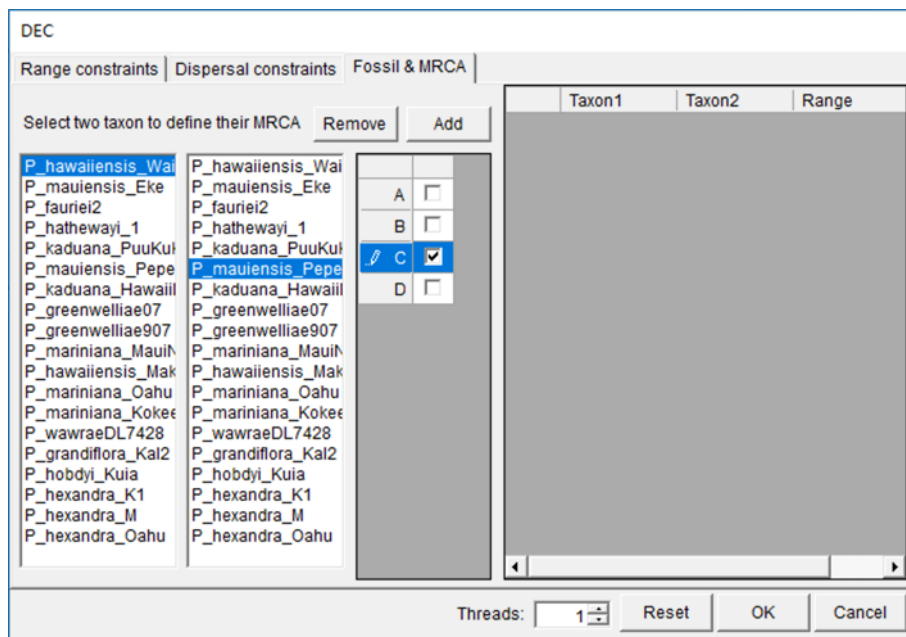
Select two species to define a clade and then specify the ancestral distribution of the clade. This is used when wanting ancestral state results for only particular nodes (for instance, there are fossils on the node).

3.3.3 Bayesian inference for discrete Areas (BayArea)

Choose [**Reconstruction >On Condensed Tree >Bayesian inference for discrete Areas (BayArea)**] to use
Bayes-Lagrange model
Note: Bayarea Analysis will deal with CONDENSED TREE ONLY

If you have a Geographic data file, click **[Load]** button, navigate to your file and select it.
[Example: Open "[RASP folder]\Sample\ Psychotria\RASP\coordinates.csv" in folder Sample]
You can also input and revise the coordinates in the entry fields in the Geographic data column.
The following settings are described in BayArea's manual (Landis, 2013):
Chain length: Number of MCMC cycles.
Frequent of sample: The MCMC sample frequency to populate area states and probabilities.
Guess Initial Rates: Use heuristic to initialize MCMC area loss/gain rates (Landis, 2013).
Geo Distance Power: Constrains the distance power parameter to be positive when True. This may lead to more sensible
parameter inferences when range data are sparse and global.
Geo Distance Truncate: When True, the pairwise distance calculator approximates extremely improbable dispersal events

as having a very small positive rate (effectively zero). This introduces negligible error into the analysis, but greatly speeds up computations when the number of areas is large.
Save original result to: save BayArea output files to another place. To visualize your MCMC results, we recommend using the ".parameters.txt" file with Tracer (http://tree.bio.ed.ac.uk/software/tracer/). After the MCMC cycle is completed, a "tracer view" will show. Before the calculation of the posterior, the chain must first burn-in. A small Burn-in value will give inaccurate results
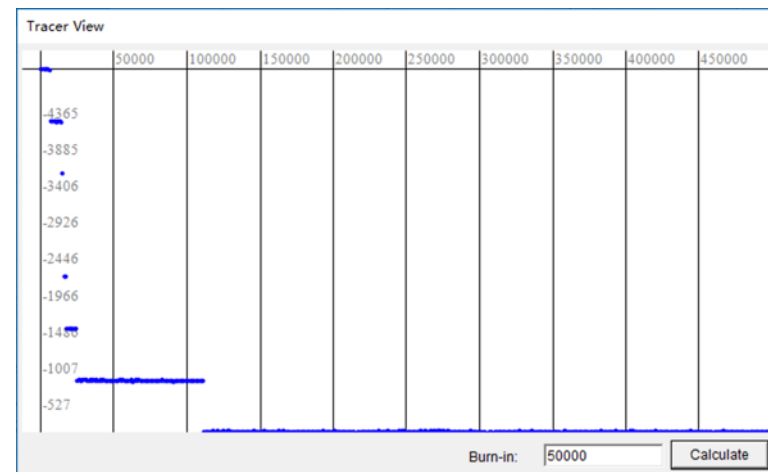
3.3.4 DIVALIKE/DEC/BAYAREALIKE in BioGeoBEARS

Note: We use biogeography of the Hawaiian shrub Psychotria, as a sample dataset. You could load the data from "[RASP folder]\Sample\Psychotria\"

Choose [**Reconstruction > On Condensed Tree > DIVALIKE/DEC/BAYAREALIKE in BioGeoBEARS**] to show this:

Include -> and **Exclude-<**: Include or exclude ranges from calculation. This can be accomplished by selecting ranges from the Include list and clicking the arrow to move them to exclude. When you do this, only the selected ranges will be excluded. You can also exclude all ranges that include a particular subset of areas using the range matrix. For example, unchecking the box in the first row, second column will remove all ranges that include A and B; AB and ABC in the example above.

Max areas: The number of unit areas allowed in ancestral distributions.

Note: In BioGeoBEARS, the value of max areas could no less than the maximum number of unit areas of species. For instance, if one species distributed in ABC, then the max areas should >= 3.

Note: Choose your favor model in the selection box at lower left.

Cores: You use more cores of CPU to speed up (Only in windows)

Note: A command window of R will display after you click OK. Please do **NOT** close it.

The table of the form designates what time slices or periods will be considered. This is important when designating the above matrix if different connectivities/areas/distance are used for different times, and it can also be important when estimating the rate matrix parameters if you want different estimates for different time periods. You could add or remove the periods from the table using the "**Add**" or "**Remove**" button. The left part of the form contains the rate matrix connectivity parameterization.

For instance, the left matrix says that for the period 1 to 2 (In time table, from 0.5 to 1.9) there is less dispersal between areas D and A. And in the period from 0 to 1, all dispersalrates are equal to 1.
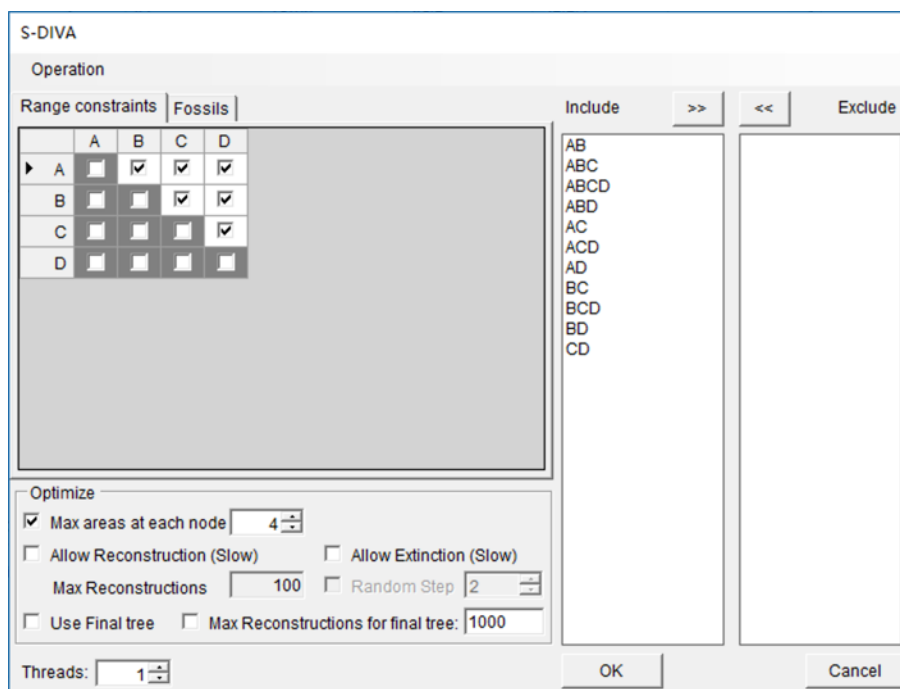
You could export/import the matrix to/from file.

[Example: Open "[RASP folder]\Sample\Psychotria\timeperiods.txt"]

3.3.5 Statistical Dispersal-Vicariance Analysis (S-DIVA)

Note: S-DIVA Analysis will deal with BOTH CONDENSED TREE AND TREES DATESET

Choose [**Reconstruction > On Trees >Statistical Dispersal Vicariance Analysis (S-DIVA)**] to use default options to run Analysis.

[Options]

Include -> and **Exclude-<**: Include or exclude ranges from calculation. This can be accomplished by selecting ranges from the Include list and clicking the arrow to move them to exclude. When you do this, only the selected ranges will be excluded. You can also exclude all ranges that include a particular subset of areas using the range matrix. For example, unchecking the box in the first row, second column will remove all ranges that include A and B; AB and ABC in the example above. Once the desired boxes have been unchecked, [**Operation> Refresh the Range List**].

We modified the source code of DIVA (Ronquist, 2011) to exclude impossible ranges.

Max areas at each node: The number of unit areas allowed in ancestral distributions.

Allow Extinction (Slow): Traditionally, there are four different biogeographic events under the event model in DIVA: dispersal, extinction, vicariance and duplication (Ronquist, 1997). In original DIVA algorithm, there is no need for calculating (Ronquist, 2001) the cost of extinction events. if some user-specified ranges are excluded, a null (or empty) result may occur; namely if the only geographic ranges that are consistent with the rule have been eliminated. For example, suppose that the total distribution is {A, B, C}, N_L = A and N_R = B. When geographic range AB is excluded, ABC should be proposed as the ancestral range, but ABC violates the rule set of the DIVA algorithm. Therefore, we have made the following modification in S-DIVA: Assume that the ancestral range of the node, i, is A_i, then the descendant nodes (terminal) are N_L and N_R. Let $|X|$ be the number of elements in X. Then the cost of an extinction event E_i could be calculated as $|A_i| - |N_L \cup N_R \cap A_i|$. When no ranges are excluded, the algorithm of S-DIVA is as same as DIVA.

Allow Reconstruction: Unchecking this option changes the method used for calculating $F(x_n)$ from i/D_t to $1/N$. Also see Yu et al. (2010) and Harris & Xiang (2009) in which the differences between these options are discussed.

Max Reconstructions: Total number of reconstructions will be kept for each tree in trees dataset.

Random Reconstructions: Save random reconstructions. We strongly recommend kept this option checked or it will cost a very long time and a large disk space to do the analysis.

Max Reconstructions for final tree: Total number of reconstructions will be kept for final tree.

Threads: How many threads will be used in analysis.

3.3.6 Statistical Dispersal-Extinction-Cladogenesis (S-DEC)

Choose [**Reconstruction >On Trees >Statistical Dispersal-Extinction-Cladogenesis (S-DEC)**] to use Bayes-Lagrange model

Note: This Analysis will deal with BOTH CONDENSED TREE AND TREES DATESET

This method is previous called Bayes-LAGRANGE. Bayes-LAGRANGE approach it is possible to optimize on multiple trees whereby topological uncertainty is taken into account. RASP get the likelihood values of all possible ancestral distributions in Lagrange, and relied on a composite Akaike weight to summarize the biogeographic reconstructions across trees (Beaulieu et al., 2013). The settings are as same as above. Bays-Lagrange is not really Bayesian, but more like pseudo-Bayesian. I prefer to call it as S-DEC model.

Note: The setting of S-DEC is as same as 3.3.2. Dispersal-Extinction-Cladogenesis (DEC).

3.3.7 Statistical DIVALIKE/DEC/BAYAREALIKE in BioGeoBEARS

Note: We use biogeography of the Hawaiian shrub Psychotria, as a sample dataset. You could load the data from "[RASP folder]\Sample\Psychotria"

In statistical method, the probability (p) of an ancestral range x at node n on the final tree is calculated as $P(x_n) = (\sum_{t=1}^{m} F(x_n)_t)/m$. Where t is the selected tree, m is the total number of sampled trees, $F(x_n)_t$ is the occurrence of an ancestral range x at node n for tree t.

Note: The setting of Statistical DIVALIKE/DEC/BAYAREALIKE in BioGeoBEARS is as same as 3.3.4

3.3.8 Bayesian Binary MCMC (BBM) Method For Ancestral State

Note: BBM Analysis will deal with CONDENSED TREE ONLY

Choose [**Reconstruction > Traits Evolution >Bayesian Binary MCMC (BBM)**] to use default options to run Analysis.
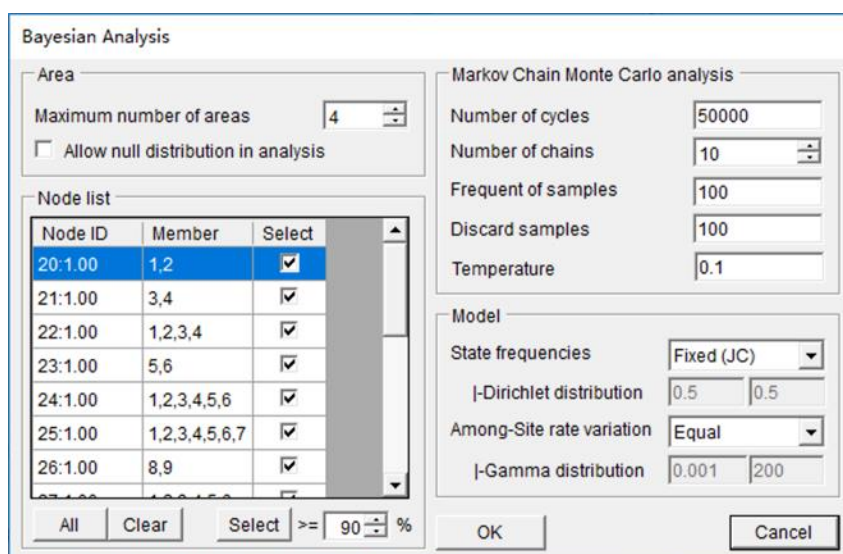
[Area]

Maximum number of areas: The number of unit areas allowed in ancestral distributions.

Allow null distribution in analysis: If the null distribution was allowed.

[Markov Chain Monte Carlo analysis]

BBM is primarily designed for reconstructing ancestral state of given nodes. It calculates the probabilities of ancestral ranges using the probabilities of each unit area generated by MrBayes. In RASP, we use the source code of MrBayes 3.1.2 (Ronquist & Huelsenbeck, 2003) to build the module of Bayesian biogeographic inference. Like BayArea, binary character states are used to code geographic range in BBM. Assume that $D = \{X_1, X_2 \dots X_n\}$ is the set of all unit areas and that $R = \{Y_1, Y_2 \dots Y_{2^n}\}$ is all possible combinations of unit areas; in other words, all possible ranges. The terms $P_1(X_i)$ and $P_0(X_i)$ represent the average probability of presence (1) and absence (0) over all sampled generations generated by MrBayes, respectively, of the ancestral species in area X_i. The probability of an ancestral range in set R is calculated as:

$$P(Y_i) = \left(\prod_{X_i \in Y_i} P_1(X_j)\right)\left(\prod_{X_i \in (D \cap Y_i)^C} P_0(X_k)\right)$$

BBM does not consider the branch length and support value of the tree, and itaccepts trees with polytomies directly giving the researcher more flexibility in analysis. Researchers also have the ability to collapsebranches that they do not wish to utilize (i.e. branches with low support value) in BBM. In BBM, "null range" (an ancestral range contains none of the unit areas) could be removed after the MrBayes analysis. The null range is useful in ancestral state reconstruction (e.g., flower spots; flowers with red spots are A, with yellow spots are B, then without spots have the "null range"), but does not make any sense in biogeographic analysis.

Number of cycles (**Ngen** in Mrbayes): This option sets the number of cycles for the MCMC algorithm.

Number of chains (**Nchains** in Mrbayes): How many chains are run for each analysis for the MCMC variant.

Frequent of samples (**Samplefreq**in Mrbayes): This specifies how often the chain is sampled.

Discard samples (**burnin** in Mrbayes): This option determines how many samples will be discarded before calculating the partition frequencies.

Temperature (**Temp** in Mrbayes): The temperature parameter for heating the chains.

Node List: In Bayesian Binary MCMC Analysis, you can deselect nodes to exclude them from the analysis. Using the Select > n% options, you can select all nodes supported by posterior probability (pp) greater than your specified value.

Note: In order to select nodes with greater than n% pp support,

[Model]

State frequencies: Character state frequencies.

Among-site rate variation: This parameter specifies the prior for the gamma shape parameter for among-site rate variation.

3.3.9 MultiState Reconstruction in BayesTraits

RASP 4.0 support multistate model in BayesTraits using ML or MCMC. After the commands are ready, please click "Run" to run BayesTraits. When analysis complete, a tree view window and state view will be shown. You could also analysis the *.log.txt file from [View->State View] in the main window:
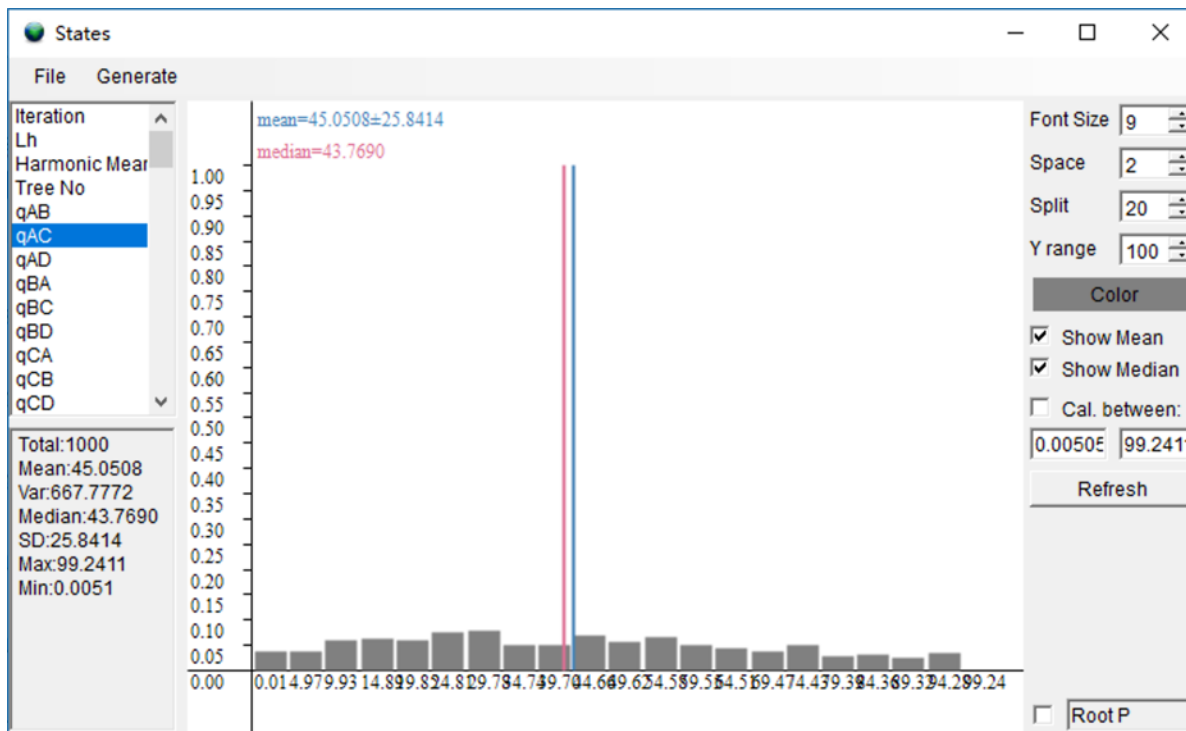
3.3.9 ER/SYM/ARD in APE

This method is implemented in 'ace' function in R package 'APE' (Paradis and Schliep 2018)

3.4 Comparison

3.4.1 Trees vs. Trees

Using [RASP folder\Sample\ Simulation\ trees_dataset.trees] as example.

Choose [**Comparion > Trees vs. Trees**]

In the new opened window, choose **[Analysis -> Calculate Matrix]** to select a method to calculate tree distance matrix.

The triplet distance (Critchlow, et al. 1996) enumerates all subsets of leaves of size three and count how often the topologies induced by the three leaves agree in the two trees. The triplet distance is intended for rooted trees, where the triplet topology is the smallest informative subtree (Sand, et al. 2014).

The SPR distance (de Oliveira Martins, et al. 2008; De Oliveira Martins, et al. 2014) is also related to the minimum number of recombination events that took place between two trees, and it designed for handling large topologies with a limited number or type of recombination.

The Robinson-Foulds distance (Robinson and Foulds 1981) is the most widely used topology metrics, and it is based on the split decompositions of the two topologies and is simply the number of edges that have no counterpart in the other topology. The length of a path from one tip to another in a tree is the number of edges within the tree that must be crossed to navigate from one tip to the other. The path difference metric (Steel & Penny, 1993) is the square root of the sum of squares of each of these differences.

The KF distance (Kuhner and Felsenstein 1994) is the sum of squares of the differences between each branch's length in the true and deduced trees, which is sensitive to the correctness of branch lengths.
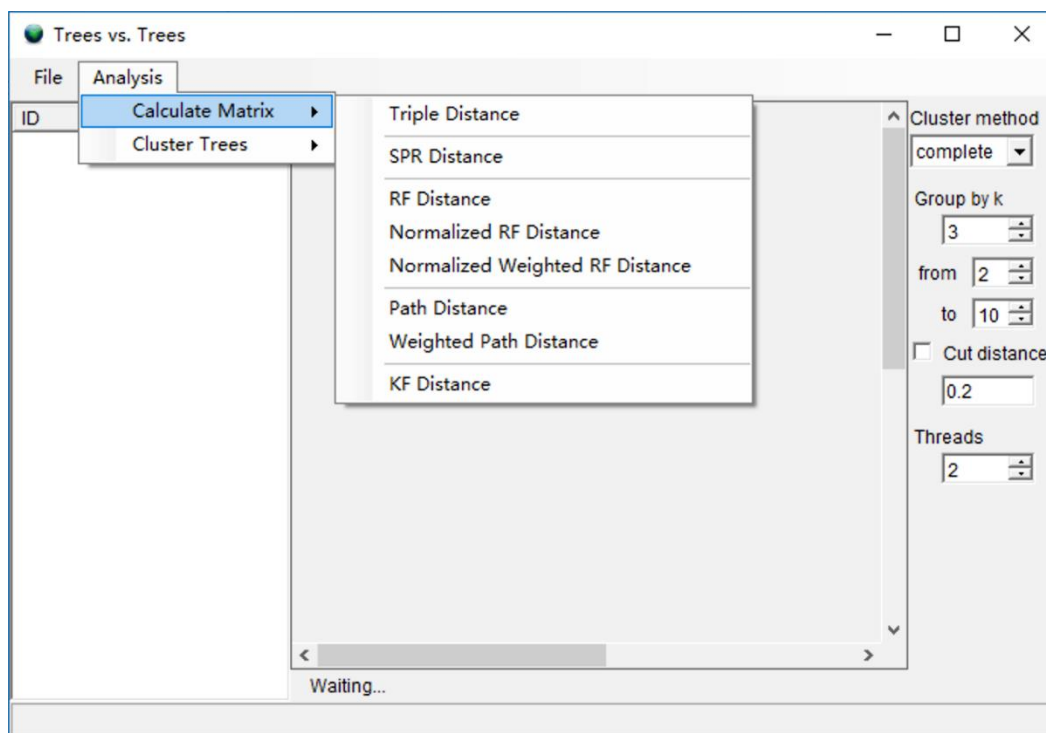
After the process completed, choose **[Analysis -> Cluster Tree -> Test k]** to test the best number of group (k value).
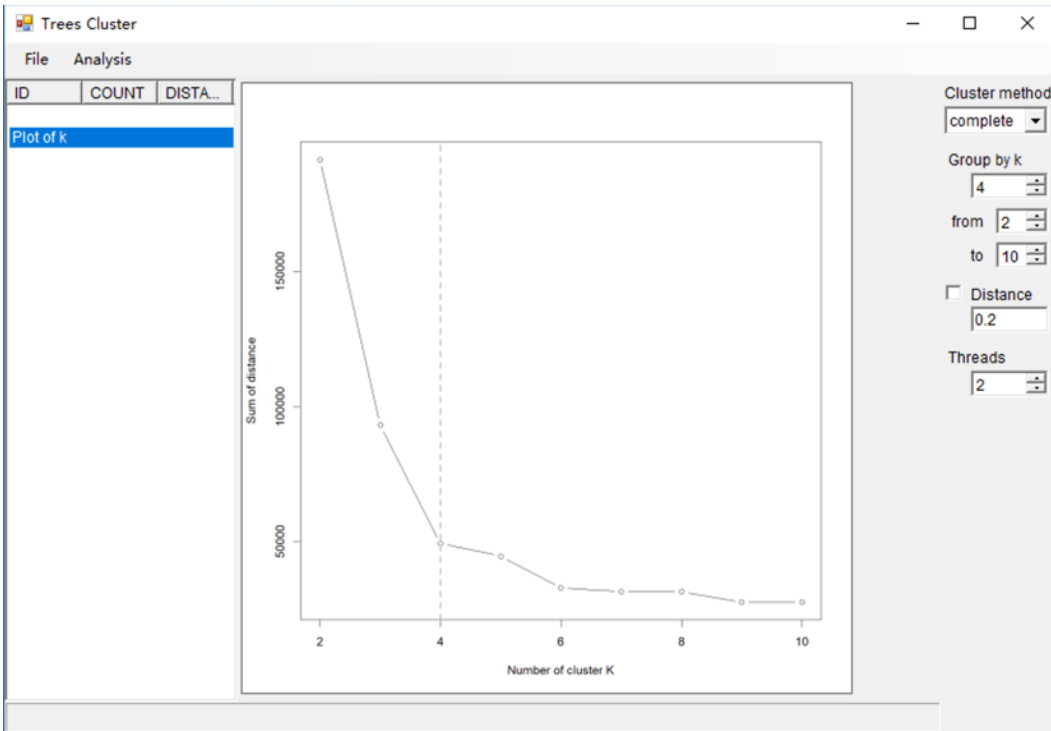
Choose [**Analysis -> Cluster Tree -> Hierarchical Cluster**] to cluster the tree using the k value.
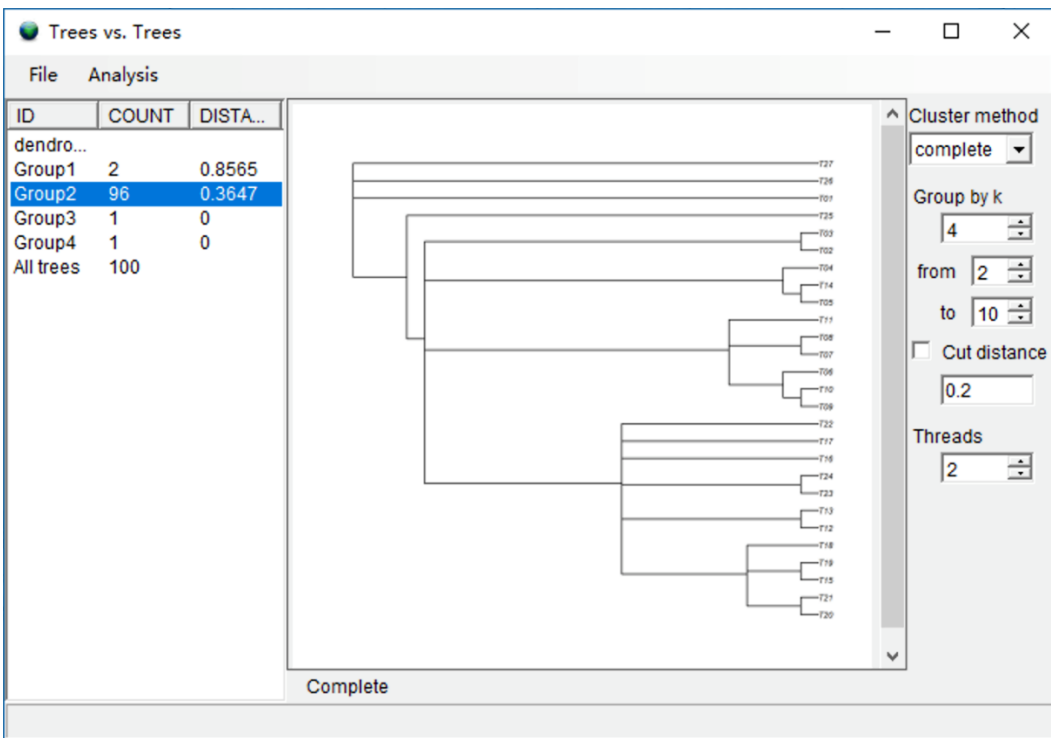
[Left list]

ID: the group id of cluster

Count: the number of trees of each group

Distance: the average distance among the group of trees.

[Note: You could double click the group id to view the condensed tree of the selected group]

[Option]

Cluster method: choose the cluster method.

Group by k: setting the number of clusters

From … to … : The range of k when using [Analysis -> Cluster Tree -> Test k]

Cut distance: Change the distance larger the given value to 1

Threads: How many threads will be used in analysis.

Having the distance matrix, the gene trees was then cluster into groups use hierarchical cluster analysis in R package (R Core Team 2018). The end result may provide insight into the sources of the heterogeneity among gene histories, for instance the genes are found to share the same history may have relationship with particular character. Additionally, this method could also be used to provide a candidate species tree, by compare the number of genes in each cluster.

3.4.2 Trees and States

Choose **[Comparion > Trees and States]**

To test for phylogenetic signal for continuous states, RASP calculates Moran's I (Moran 1948, 1950), Abouheif's C_{mean} (Abouheif 1999), Pagel's λ (Pagel 1999) and Blomberg's K (Blomberg, et al. 2003) using the R package 'adephylo' (Jombart, et al. 2010). For discrete states, RASP fits models of trait evolution using a likelihood ratio test and calculates Pagel's λ using the R package 'geiger' (Pennell, et al. 2014). These methods could be applied for all characters on a single tree (using the Tree vs. States tool) or on a single character with all input trees (using the State vs. Trees tool).

Tree and States — □ ✕

File   Analysis

ID	State	Tree	Moran's I	P-value (I)	Cmean	P-value (C)	Pagel's λ	P-value (λ)	K	P-value (K)	D count
1	Contin...	((((((((1...	-0.027657	0.207	0.093998	0.139	0.722293	0.028488	1.256612	0.002002	-
2	Discrete	((((((((1...	-	-	-	-	1	0.003904	-	-	8
3	Multist...	((((((((1...	-	-	-	-	-	-	-	-	17

For research who interested in the relationship between gene tree and geographical distribution, we propose a simple method to measure the fit between a tree and geographical distribution. Assume the species is distributed in area M, and that all possible subclades of the tree is N. Let A_x be the set of taxa distributed in area x (x ∈ M), S_j be the set of taxa of subclade j (j ∈ N), |X| be the number of elements in X, the total cost is: $D_{count} = \sum_{x \in M} \min_{j \in N}(|A_x| + |S_j| - 2 * |A_i \cap S_j|)$. For a given distribution or a given tree, smaller D_{count} values indicate a better fit between a tree and the geographic distribution.

3.5 Results

3.5.1. Save and Load Results

After running an analysis, you can save the last Result from **[File > Save > Result]**. The result file could be load into RASP from **[File > Load Result]**. To combine several results, see part 4.4.

3.5.2 View results on tree

After running an analysis, you can access the **Tree View** window from [**Graphic > Tree View**] to see the graphic result.
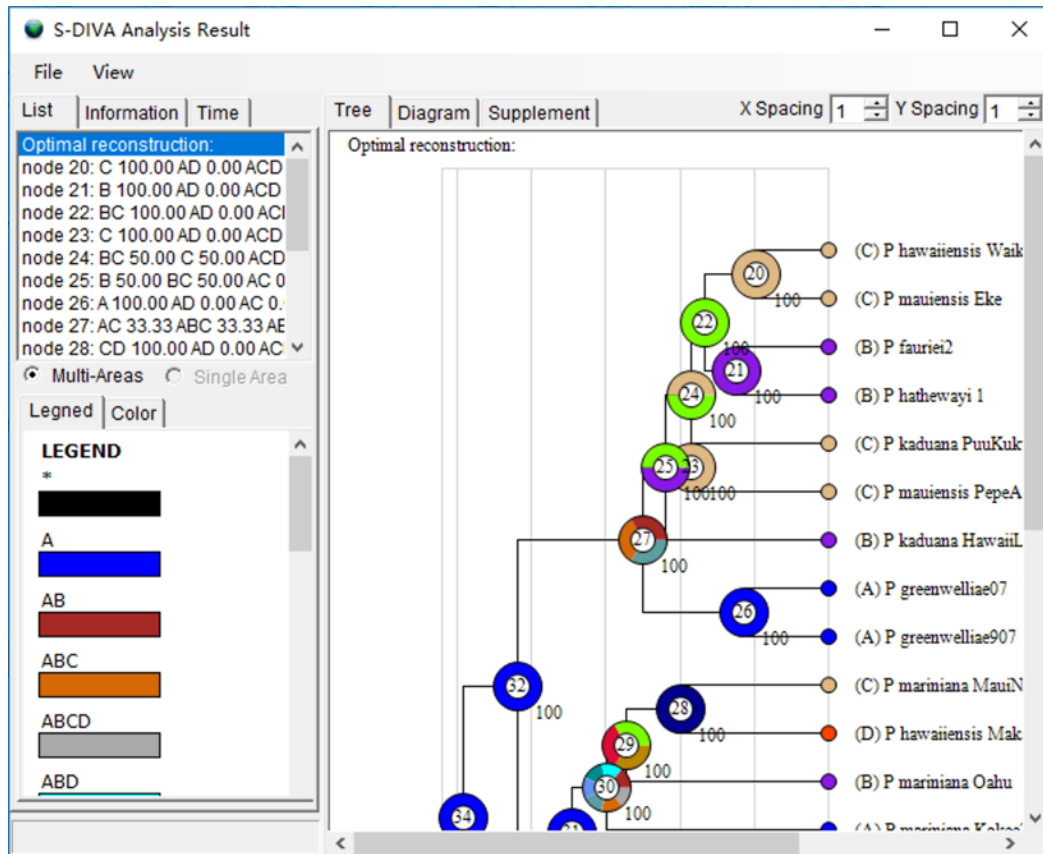
[Note: S-DIVA, Lagrange and Bayes-Lagrange only have "Multi-Areas" mode, for they could not give the probabilities of each single area]
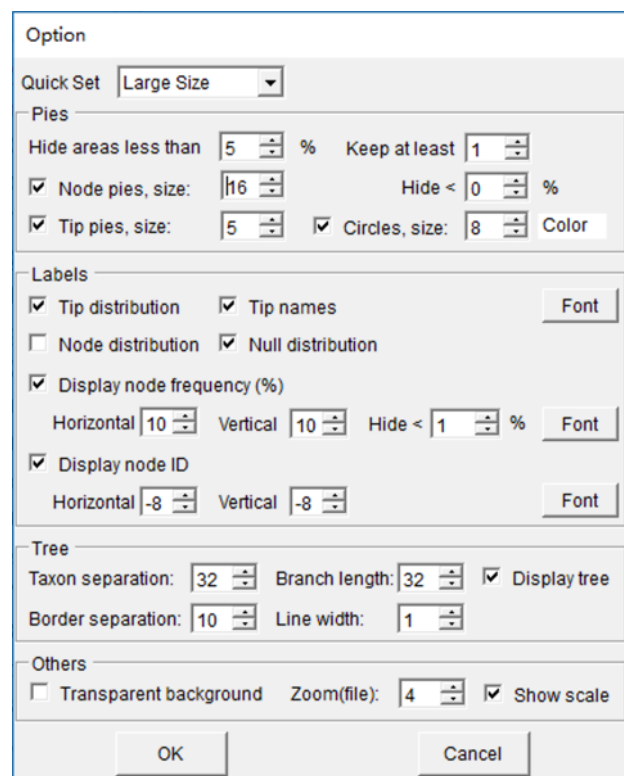
[Options]
Open **[View > Option]** in **Tree View** window to show the options.
Transparent BG: Save PNG file with a transparent background.
Quick Options: Size of the displayed tree and its annotations.
Hide areas lower than: Hide areas which probability of lower than a specific number.
Keep at least: How many areas displayed at a node at least.
Display Lines: Show lines of tree or not.
Display area distribution: Show distribution areas. These will display as colored letters next to the pie charts.
Display area pies, radii: Show pie charts of distribution area and set the radii of pie charts (5-1000).
Taxon separation: Set the separation of taxa (10-1000).
Branch length: Set the length of branches (10-1000px). All branches in the cladogram become proportionately larger or smaller.
Border separation: Set the width of the white space in the tree view window (10-1000). Note that it is possible to make this too

small for the width of the tree.
Line width: Set the width of the line (1-10)

[List] page
result of *: Select a node to display alternative ancestral distributions (pie chart and bar chart) at the selected node on thetree.
Muti Areas, Single Area: two different kind of distribution range pie. TheSingle Area only supports result of BBM and BayAreamethod. It shows the probabilities of presence of each single area.

[Information] page
Applying the DIVA model (Ronquist, 1997) to S-DIVA results can be challenging because the former does not accept trees with polytomies. Therefore, we have expanded the DIVA event model to a generalized event model, DEV (Dispersal, Extinction and Vicariance), which accepts polytomies. (When the tree does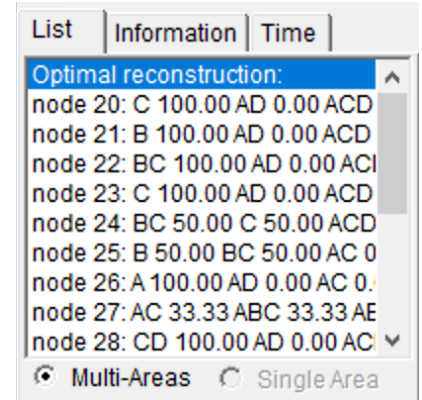 not contain polytomies, DEV model is the same with DIVA model). Traditionally, there are four different biogeographic events under the event model in DIVA: dispersal, extinction, vicariance and duplication (Ronquist, 2001). In the DEV model, we define only three biogeographic events: dispersal, extinction, and vicariance. We assume that both dispersal and duplication result from biological processes inducing or leading to range expansion. This differs from vicariance, in which abiotic processes subdivide an ancestral geographic range, and from local extinction, which results in range contraction.

To apply the DEV model to biogeographic reconstructions, we consider only the most likely state (MLS) of nodes. Assume that the MLS of the ancestral node, i, is A_i, then the descendant nodes (terminal) are $N_1, N_2 \ldots N_j$, where j is the total number of descendant nodes. We define $N_u = \bigcup_{k=1}^{j} N_k$, $N_s = \bigcap_{k=1}^{j} N_k$. Let $|X|$ be the number of elements in X. The variables D_i, E_i and V_i are the dispersal, extinction and vicariant events of the node i respectively. One can show that:

$$D_i = \begin{cases} |N_u \cup A_i| - |A_i| & , |A_i| > 1 \\ |N_u \cup A_i| - i * |N_s \cap A_i| + \sum_{k=1}^{j}(|A_i \cap N_k|) & , |A_i| = 1 \end{cases}$$
$$E_i = |A_i| - |N_u \cap A_i| \qquad V_i = \begin{cases} i - 1 & , |N_s \cap A_i| = 0 \\ 0 & , |N_s \cap A_i| > 0 \end{cases}$$

The DEV event model assumes that the MLS at each node are the true state.
[Note: the DEV model may misinterpret the actual processes assumed by DEC, BayArea and BBM, as these methods do not minimize the number of events using the parsimony process.]
Dispersal: Highlight the nodes that have Dispersal event.
Vicariance: Highlight the nodes that have Vicariance event.
Extinction: Highlight the nodes that have Extinction event.

[Time] page
Reverse: Convert the branch length from before to nowadays.
Time of root node: Set the time of the root of the tree.
Calculate: Re-Calculate the time of each node using the time of the root and calculate the time-event curve.
In the biogeographical analyse, researchers often interested in the relationship between the nodes and biogeographical events on dated phylogenetic tree. However, it might be very difficult to do the analysis on a large and complex tree, especially when the molecular dating is not very accurate. The extent of influence of geographical event on phylogenetic tree is hard to define, and the qualitative and quantitative analyses are not easy to make. In RASP, we develop a time-events curve to analysis the dated phylogenetic tree. The events on the node were treated using a modified Gaussian distribution. Let n be the total number of tree, T_i is the time of node i. One can show that, u is the unit time of the events:

$$D(t) = \sum_{i=1}^{n} D_i \left(e^{-\frac{\left(\frac{t-T_i}{u}\right)^2}{2}} \right) \quad E(t) = \sum_{i=1}^{n} E_i \left(e^{-\frac{\left(\frac{t-T_i}{u}\right)^2}{2}} \right)$$

$$V(t) = \sum_{i=1}^{n} V_i \left(e^{-\frac{\left(\frac{t-T_i}{u}\right)^2}{2}} \right) \quad S(t) = \sum_{i=1}^{n} \left(e^{-\frac{\left(\frac{t-T_i}{u}\right)^2}{2}} \right)$$

$$M(t) = D(t) + E(t) + V(t)$$

In these four functions, D(t), E(t),V(t) and M(t) implie the duplication, extinction, vicariance event and the total number of biogeographic events through the time respectively. S(t) implies the number of nodes through the time. Noting that the unit time of event has a great influence on the shape of time-events curve. One could find an appropriate value according to the duration of geological events.
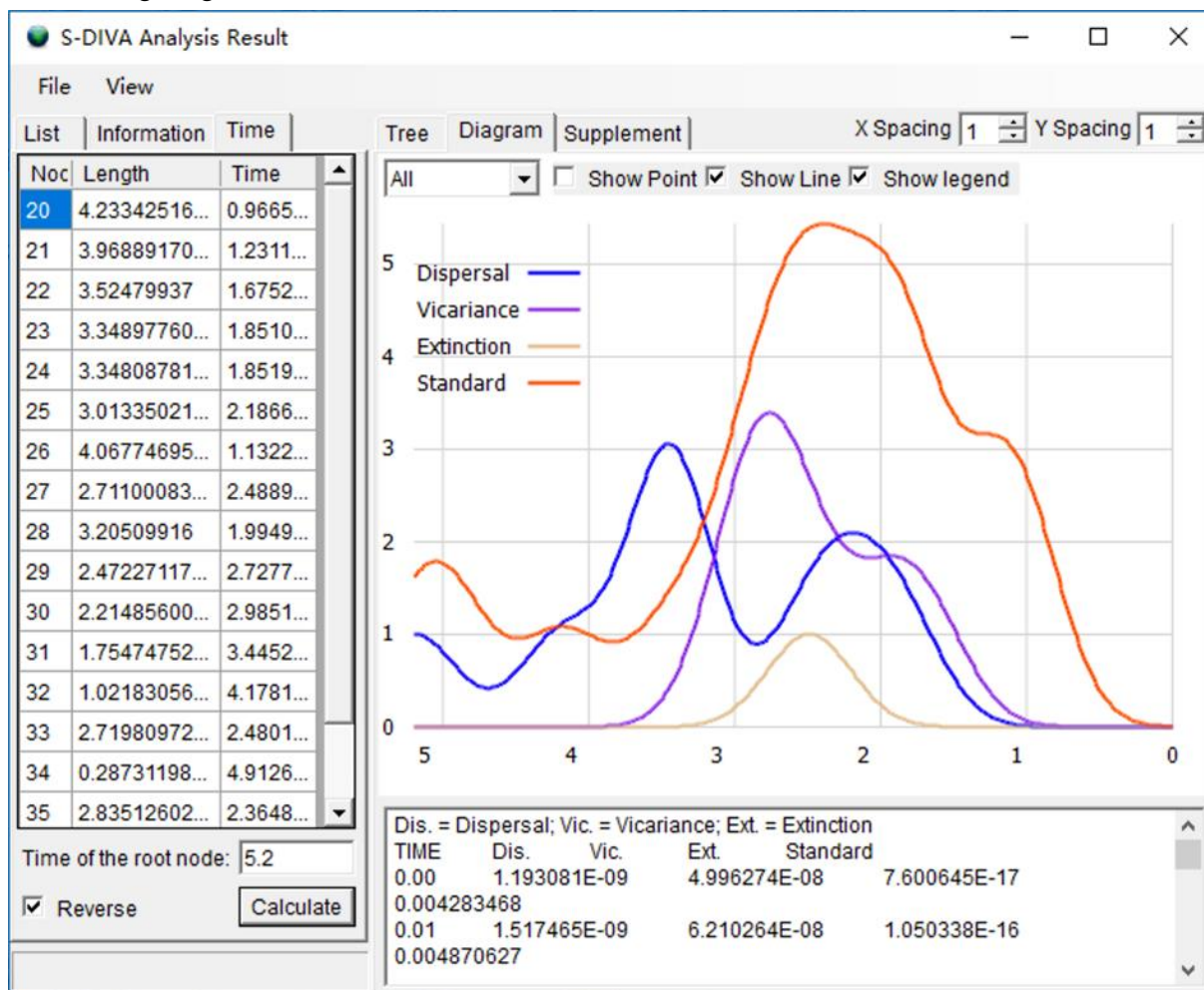
The graphic could be saved from [File->Export Graphic->Export Diagram]

[Mouse control]

[*Double click* on a node]: Select the node.

[*Double click* on blank space]: Cancel node select.

[Save/Loadresults]

Open **[File -> Save Result]** to save the text result.

Open **[File -> Save Info]** to save the text in Diagram page.

Open **[File -> Save Current Tree]** to save the tree in Phylip format.

Open **[File ->Export Graphic ->*]** to save your tree/legend/ dtagram as graphic file.

Open [**File > Load Result**] and navigate to your saved text result and select it.

3.5.3 Node view

When "Estimate for nodes" in S-DIVA analysis is enabled, you could view the results of nodes. You could also view the results of nodes in BBM method. The left figureshows a result for estimating a particular node. This can be accessed from [**Graphic > Node View**].

Hide areas lower than: Hide areas which probability of lower than a specific number.

Note: Hidden probabilities are lumped together and reported as * (see far right bar in the figure to the left). In this example, ranges with probabilities less than 5% are hidden.

Keep at least: Fewest number of areas displayed (trumps the Hide areas lower than setting).

[**Save results**]

Open [**File -> Save Result**] to save the text result.
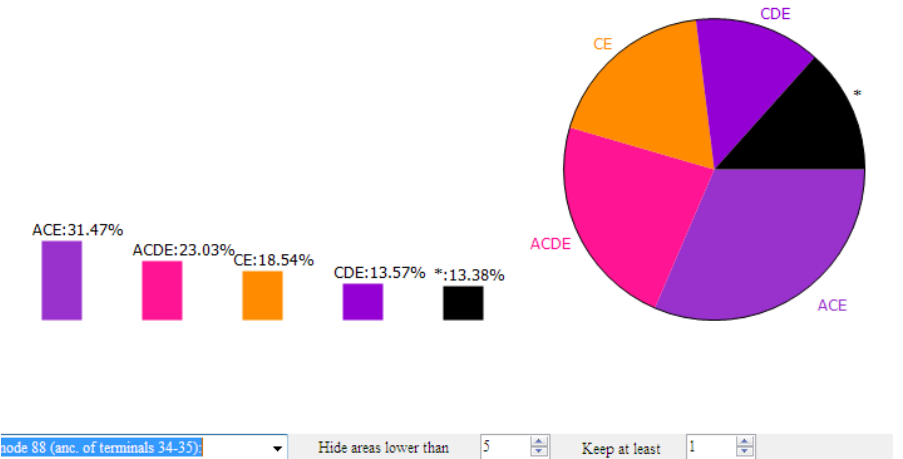
3.5.4 Tracer View

View likelihood value of BayArea MCMC cycle and re-calculate the result with the new burn-in value.

This can be accessed from [**Graphic > Tracer View**]
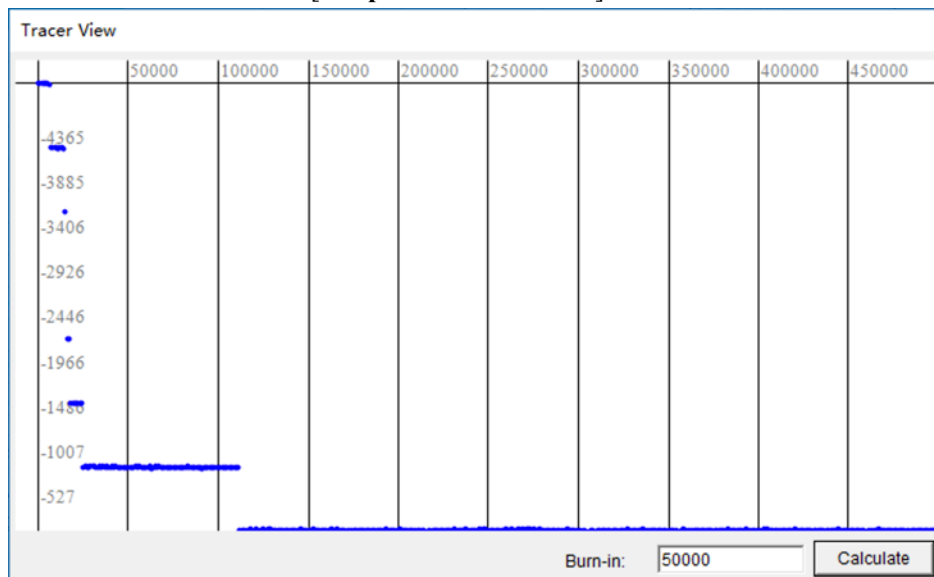
4. How to …

All sample files referenced in this manual are in the sample folder.

4.1 How to make Trees data set.

4.1.1 How to obtain trees data set from BEAST. (Back to Top)
- Launch **BEAUti v*.exe**
- Select **File > Import Alignment** and navigate to your NEXUS input file.
  (From our sample files, we would select *sample1.nex* in the Sample folder)
  (Note: If **BEAUti** can't load the NEXUS file then load the file into PAUP and export it use "**format=NEXUS**" option.)
- Select MCMC panel, set the number of generations the MCMC algorithm will run for.
  (We set **length of chain** = *1000000* to do a quick run)
- Click **Generate BEAST file…** and save your file.
  (We saved it as *samlpe1.xml* in *sample* folder)

- Launch **BEAST v*.exe**
- Enter a **Random number seed** like 12345
- Choose your BEAST XML input file.
  (From our sample files, we would select *sample1.nex* in the Sample folder)
- **Run** it!
- After the program is finished, you will find a *.trees* file in the same folder of your.xml file.
  (In our example, *Sample1.trees*is our trees data set.)

4.1.2 How to obtain trees data set from PAUP. (Back to Top)
- Use **Lset** or **Pset** command set for ML or MP analysis with option "**Collapse= NO;**"
- Define outgroups and root the trees with **"roottrees OUTROOT=MONOPHYL;"**
- Save all of your trees using "**format=NEXUS**" option.

4.1.3   How to obtain trees data set from other phylogenetic programs. (Back to Top)
If you are using Mrbayes, it is helpful to define outgroups and specify rooting before you run mcmc. If you did not define outgroups and specify rooting before the MrBayes run, you can root all trees using PAUP or other available software. You can load the MrBayes output file (**.run.t*) as your trees data set once the trees have been rooted.
For other phylogenetic programs, there are two methods for making a trees data set:
Method 1:
- Save the trees as Nexus format from whatever phylogenetic program you are using.
Method 2:
- Save the trees as PHYLIP format.
- Load the trees into PAUP then export the trees use "**format=NEXUS**" option.

4.1.4   *The bare essentials for the tree file*. ([Back to Top](#))

Our sample trees file includes considerable amounts of phylogenetic program
output, but here is what a trees file **must** contain. An example, which can be
accepted by RASP, is shown to the left.

- Your taxa should have unique labels or names (upper arrow).
- Translate your unique taxon names into integers (1- ~) (middle arrow).
  These integers are required for the ID field in the RASP program.
- Your file should contain one or more phylogenetic trees in set notation
  (lower arrow). Many tree generating programs such as PAUP*,
  Mrbayes, etc. include a lot of extra information within each tree such
  as branch lengths, likelihood values, and so on. This extra information
  does not hinder RASP, but it is also not essential.

Note: Some programs used to root trees, place the outgroup at the beginning
on the tree strong (like the outgroup 3 in the example), others do the
opposite, placing the outgroup at the end of the strong. RASP can accept
both types of rooting.

```
Begin taxa;
Taxlabels   ←
Taxon_1
Taxon_2
Taxon_3
Taxon_4
Taxon_5
Taxon_6
Taxon_7
Taxon_8
Taxon_9
Taxon_10
;
end;

Begin trees;
Translate   ←
1 Taxon_1,
2 Taxon_2,
3 Taxon_3,
4 Taxon_4,
5 Taxon_5,
6 Taxon_6,
7 Taxon_7,
8 Taxon_8,
9 Taxon_9,
10 Taxon_10
;

tree 1 = (3,((8,1),(6,(((2,(4,9)),5),(10,7)))));
tree 2 = (3,(((((5,(6,2)),7),(4,9)),10),(1,8)));   ←
tree 3 = (3,((((2,6),(((4,9),5),(7,10))),8),1));
tree 4 = (3,(5,(((((8,1),(9,4)),10),7),(2,6))));
tree 5 = (3,((((4,(10,(((1,8),6),2))),9),7),5));

end;
```

4.2 How to make a Condensed tree.

4.2.1 How to make a condensed tree by Tree Annotator. ([Back to Top](#))

- Launch *TreeAnnotator*.exe*
- Set **Burnin** and **Posterior Probability**
- Choose your trees data set as Input Tree File
  (In our example, we chose *sample\sample1.trees*)
- Choose Output File
  (We saved it as *Sample1_Final_Tree.tre* in *sample* folder. *Sample1_Final_Tree.tre* is our condensed tree.)

4.2.2 How to make a condensed treeusing PAUP. ([Back to Top](#))

- Define outgroups androot the trees with **"roottrees OUTROOT=MONOPHYL;"**
- Export a tree using "**savetrees format=NEXUS**" commands.

4.2.3 How to obtain a condensed tree from other phylogenetic programs. ([Back to Top](#))

If you are using Mrbayes, you should [make the condensed tree using Tree Annotatoror or PAUP*.](#)
To make a condensed tree using other phylogenetic programs, there are three methods for making a condensed tree:
Method 1:
- Save the tree as Nexus format from whatever phylogenetic program you are using.

Method 2:
- Save the tree as PHYLIP format.
- Load the tree into PAUP then export the tree use "**format=NEXUS**" option.

Method 3:
- Make a text tree file by yourself as the following format:
  tree tree1=(((((6,((2,4),((9,8),3))),1),(7,11)),10),5);
  or
  (((((6,((2,4),((9,8),3))),1),(7,11)),10),5);
  Important: Trees with polytomies could only be used in Bayesian Binary MCMC and MP analysis!

4.2.4The bare essentials of the condensed tree. (Back to Top)

Condensed tree files may contain various types of commands and information relevant to the program that created them (e.g., #NEXUS, Begin trees;, etc.) An RASP condensed tree *must* contain a line of text that looks like:

Tree tree1=(((((6,((2,4),((9,8),3))),1),(7,11)),10),5);

Or simply:

(((((6,((2,4),((9,8),3))),1),(7,11)),10),5);

Taxon names must be replaced with integers as shown. The unique integer representing each taxon should be the same as in the trees file. Other types of information in the condensed tree file may not hinder RASP but are not required.

4.3 How to make Distributions file

4.3.1 How to make a Distributions file (Back to Top)

Name the distributions **A, B, C, etc.** and specify multiple-area distributions like **BD or ACE**. Only letters from **A to O**can be used (Ronquist, 1997, 2001).

- Launch **RASP**
- Select **File > Load Trees** and navigate to your trees data set
- Select **File > Save Distribution** and save it as a .csv file
- Open the your saved .csv file in a text editor or Excel
- Input the distributions after the species name like this:

Note: RASP can tolerate differences in taxon names between the trees file and the distribution file (i.e., Species_01 in your trees file compared to Species01 or Any name in the taxon or B column). RASP above 3.x also support load and save distributions in PHYLIP format, it looks like that:

```
19      4
P_hawaiiensis_WaikamoiL1        0010
P_mauiensis_Eke 0010
P_fauriei2      0100
P_hathewayi_1   0100
P_kaduana_PuuKukuiAS    0010
P_mauiensis_PepeAS      0010
P_kaduana_HawaiiLoa     0100
P_greenwelliae07        1000
```

4.3.2 How to input distributions in RASP. (Back to Top)

Name the distributions **A, B, C, etc.** and specify multiple-area distributions like **BD or ACE**. Letters from **A to O** must be used (Ronquist, 1997, 2001).

- Launch **RASP**
- Select **File > Load Trees** and navigate to your trees data set
- Type your distributions directly into the **State** column.
- Remember to save the distributions you have entered to a file for use in future analyses. Save the distributions as a .csv file using **File > Save State**.
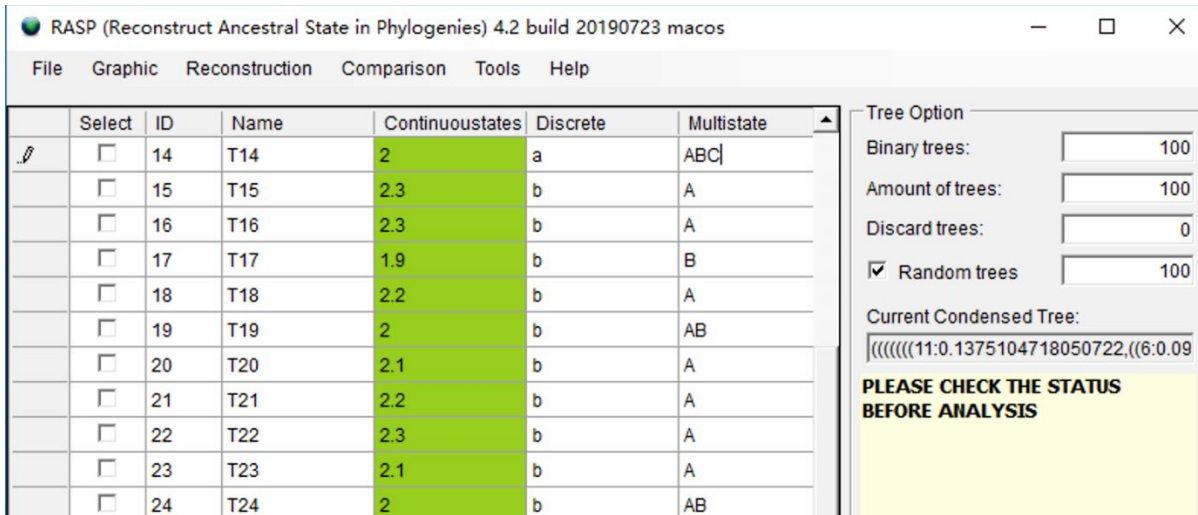
4.4 How to combine results

RASP Results generated from the same dataset with the same method could be combined. You could run a method (like Bayarea) several times to get several results, and then combine them to a single result.

Open **[Tools > Combine Results]** to combine and save different result.

4.5 How to remove Outgroups

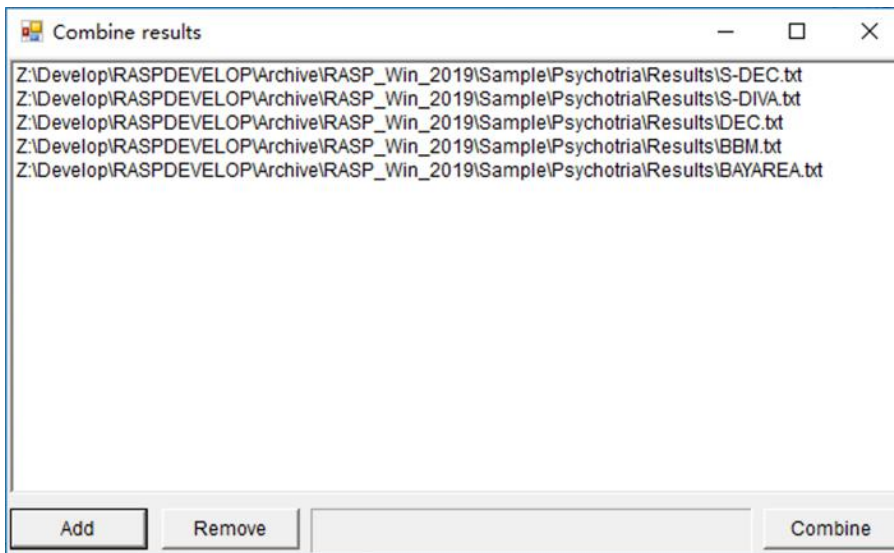Distributions of outgroups may pose limitations to historical biogeographic analysis for two reasons. First, an outgroup may represent multiple species that are widely distributed. In this case, it is expected that the ancestor of the outgroup originated in only one geographic area, but has since spread to many. Another difficulty is outgroup selection. It may sometimes be essential to select an outgroup for phylogenetic reconstruction that has a large or unknown phylogenetic distance from the ingroup. While this may be suitable for phylogeny, the geographic distributions of more closely related groups may be essential for reconstructing ancestral ranges, particularly at deep nodes. We recommend to remove all of the outgroups before biogeographic analysis.

Remove outgroups in RASP is very easy. After the trees data set and condensed tree is loaded, select all of the outgroups in the main window. Select [**Tools > Remove Outgroups**] to save a new trees data set and condensed tree without outgroups. Then you cloud close the current data [File > Close Current Data] and reload the new data into RASP.

	ID	Taxon	State	Select
	1	P_hawaiiensis...	C	☐
	2	P_mauiensis_...	C	☐
	3	P_fauriei2	B	☑
	4	P_hathewayi_1	B	☐
	5	P_kaduana_P...	C	☐
▶	6	P_mauiensis_...	C	☑
	7	P_kaduana_H...	B	☐

Tree Option

Binary trees:	100
Amount of trees:	100
Discard trees:	0
☑ Random trees	100

Current Condensed Tree:

((((((((1:0.9656850499,2:0.965685049!

5. Explanation of messages

1. **Allowing reconstructions will need a larger amount of free space on disk and will take a longer time to run. Are you sure to use it?**

   User will get this message when the number of taxa is more than 64 and reconstructions are allowed such that $F(x_n)_t = \dfrac{i}{D_t}$. Allow reconstruction will need a large amount of free space on disk and may take a long time to run.

   For example, processing 100 trees with 84 taxon (max areas of each node =4) needs about 5G free space on disk and will take about 2 hours. You may click "yes" to enable "Allow reconstructions" or click "No" to disable it.

2. **Cannot format the tree!**

   The RASP program accepts phylogenetic trees (a trees dataset and a condensed tree) generated by the programs BEAST (Drummond and Rambaut, 2006), PAUP* (Swofford, 2003) and MrBayes (Huelsenbeck& Ronquist, 2003). Other tree file formats may cause this error. The error may also occur if you are using an accepted format but have syntax errors in your file (e.g., a misplaced semicolon, a missing "end" command, etc.). Opening the file using PAUP* may help you determine if there is a syntax error, and, if so, what it is.

3. **Cannot process the computed file!**

   There is something wrong with the temp file. Please delete the folder "Temp" and restart the RASP.

4. **Cannot process the trees!**

   There is something wrong with the temp file. Please restart RASP.

5. **Distributions must be labeled using the letters A to Z only!**

   Remember that distributions must be labeled using the letters **A** to **Z** only and should include no more than 26 unit areas. RASP does not highlight rows containing unrecognized characters.

6. **Burn-in error!**

   The value of the Burn-in must greater than or equal to zero but smaller than the amount the trees.

7. **Need at least one tree!**

   You need at least one tree in the trees file to run RASP!

8. **Your group of organisms should include no more than 512 taxa!**

   Your group of organisms should include no more than 512 taxa and the distributions should include no more than 15 unit areas in S-DIVA method. You may try to use BayArea method.

9. **Please change your system's number format to English! (ex. 3.14 not 3,14)**

   The decimal symbol of your system caused this problem. Please go to "control panel->Regional and Language Options" and change your "Current format" to English. It is a bug and I will try to fix it in the next release of RASP.

10. **Distributions should be Continuous letters! Please alter area …**

    The distributions in Bayesian method must be continuous letters. For instance: Right, ABCDE; Wrong, ACDEF

11. **Your group of organisms should include no more than 512 taxa!**

    Your group of organisms should include no more than 512 taxa in All method.

6. References(back to contents)

Abouheif E. 1999. A method for testing the assumption of phylogenetic independence in comparative data. Evolutionary

Ecology Research 1:895-909.

Alexandre A., Johan A.A.N., Claes P., Isabel S. (2009) Tracing the impact of the Andean uplift on Neotropical plant

evolution. Proceedings of the National Academy of Sciences of the United States of America106 (24): 9749–9754

Beaulieu J M, Tank D C, Donoghue M J.A Southern Hemisphere origin for campanulid angiosperms, with traces of the

break-up of Gondwana.BMC evolutionary biology, 2013, 13(1): 80.

Blair C, Bryson Jr RW, Linkem CW, Lazcano D, Klicka J, McCormack JE. 2019. Cryptic diversity in the Mexican

highlands: Thousands of UCE loci help illuminate phylogenetic relationships, species limits and divergence times of

montane rattlesnakes (Viperidae: Crotalus). Mol Ecol Resour 19:349-365.

Blair C, Noonan B, Brown J, Raselimanana A, Vences M, Yoder A. 2015. Multilocus phylogenetic and geospatial analyses

illuminate diversification patterns and the biogeographic history of Malagasy endemic plated lizards (Gerrhosauridae:

Zonosaurinae). Journal of Evolutionary Biology 28:481-492.

Blomberg SP, Garland Jr T, Ives AR. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more

labile. Evolution 57:717-745.

Blomberg SP, Garland Jr T. 2002. Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods.

Journal of Evolutionary Biology 15:899-910.

Bourguignon T, Tang Q, Ho SYW, Juna F, Wang Z, Arab DA, Cameron SL, Walker J, Rentz D, Evans TA, et al. 2018.

Transoceanic Dispersal and Plate Tectonics Shaped Global Cockroach Distributions: Evidence from Mitochondrial

Phylogenomics. Mol Biol Evol 35:970-983.

Choi J, Kim S-H. 2017. A genome tree of life for the fungi kingdom. Proceedings of the National Academy of Sciences

114:9391-9396.

Critchlow DE, Pearl DK, Qian C. 1996. The triples distance for rooted bifurcating phylogenetic trees. Systematic Biology

45:323-334.

de Oliveira Martins L, Leal E, Kishino H. 2008. Phylogenetic detection of recombination with a Bayesian prior on the

distance between trees. PLoS One 3:e2651.

De Oliveira Martins L, Mallo D, Posada D. 2014. A Bayesian supertree model for genome-wide species tree reconstruction.

Systematic Biology 65:397-416.

Donoghue M.J., Smith S.A. (2004) Patterns in the assembly of the temperate forest around the Northern Hemisphere.

Philosophical Transactions of the Royal Society of London: Biology359: 1633–1644.

Drummond, A.J. and Rambaut, A. (2006) BEAST v1.4. http://beast.bio.ed.ac.uk/.

Edwards SV, Xi Z, Janke A, Faircloth BC, McCormack JE, Glenn TC, Zhong B, Wu S, Lemmon EM, Lemmon AR. 2016.

Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. Molecular

Phylogenetics and Evolution 94:447-462.

Edwards SV. 2009. Is a new and general theory of molecular systematics emerging? Evolution: International Journal of

Organic Evolution 63:1-19.

Felsenstein, J. (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of

Genetics, University of Washington, Seattle.

Han J, Jian P, Michelin K. 2006. Data Mining, Southeast Asia Edition. In: San Francisco: Elsevier Inc.

Harris AJ,Xiang Q-Y. (2009) Estimating ancestral distributions of lineages with uncertain sister groups: a statistical

approach to Dispersal-Vicariance Analysis and a case using Aesculus L. (Sapindaceae) including fossils. Journal of

Systematics and Evolution47: 349–368.

Huelsenbeck J.P., Ronquist F. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, 19,

1572–1574.

Jombart T, Balloux F, Dray S. 2010. Adephylo: new tools for investigating the phylogenetic signal in biological traits.

Bioinformatics 26:1907-1909.

Kuhner MK, Felsenstein J. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary

rates. Mol Biol Evol 11:459-468.

Lamm K.S., Redelings B.D. (2009) Reconstructing ancestral ranges in historical biogeography: properties and prospects. Journal of Systematics and Evolution, 47, 369–382.

Landis M J, Matzke N J, Moore B R, et al. 2013.Bayesian analysis of biogeography when the number of areas is large. Systematic biology 62 (6): 789-804

Liu L, Yu L, Edwards SV. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC evolutionary biology 10:302.

Liu L, Yu L, Kubatko L, Pearl DK, Edwards SV. 2009. Coalescent methods for estimating phylogenetic trees. Molecular Phylogenetics and Evolution 53:320-328.

Maddison WP. 1997. Gene trees in species trees. Systematic Biology 46:523-536.

Matzke NJ. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Systematic Biology 63:951-970.

Meade A, Pagel M. 2018. BayesTraits: a computer package for analyses of trait evolution. In: Version.

Moran PA. 1948. The interpretation of statistical maps. Journal of the Royal Statistical Society. Series B (Methodological) 10:243-251.

Moran PA. 1950. Notes on continuous stochastic phenomena. Biometrika 37:17-23.

Navaud O, Barbacci A, Taylor A, Clarkson JP, Raffaele S. 2018. Shifts in diversification rates and host jump frequencies shaped the diversity of host range among Sclerotiniaceae fungal plant pathogens. Mol Ecol 27:1309-1323.

Nylander J.A.A., Olsson U., Alström P, Sanmartín I. (2008) Accounting for phylogenetic uncertainty in biogeography: a Bayesian approach to Dispersal–icariance Analysis of the thrushes (Aves: Turdus). Systematic Biology, 57, 257–268.

Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877.

Paradis E, Schliep K. 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526-528.

Pennell MW, Eastman JM, Slater GJ, Brown JW, Uyeda JC, FitzJohn RG, Alfaro ME, Harmon LJ. 2014. geiger v2. 0: an

expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics 30:2216-2218.

R Core Team. 2018. R Foundation for Statistical Computing; Vienna, Austria: 2014. R: A language and environment for statistical computing:2013.

Robinson DF, Foulds LR. 1981. Comparison of phylogenetic trees. Mathematical biosciences 53:131-147.

Ronquist F, Huelsenbeck JP (2003) MrBayes3: Bayesian phylogenetic inference undermixed models. Bioinformatics 19:1572–1574.

Ronquist, F. (1997) Dispersal-vicariance analysis:Anew approach to the quantification of historical biogeography. Systematic Biology, 46, 195–203.

Ronquist, F. (2001) DIVA version 1.2.Computer program for MacOS and Win32.Evolutionary Biology Centre, Uppsala University.Available at http://www.ebc.uu.se/systzoo/research/diva/diva.html.

Sand A, Holt MK, Johansen J, Brodal GS, Mailund T, Pedersen CN. 2014. tqDist: a library for computing the quartet and triplet distances between binary or general trees. Bioinformatics 30:2079-2080.

Sanmartín I, Enghoff H., Ronquist F..(2001) Patterns of animal dispersal, vicariance and diversification in the Holarctic.Biological Journal of the Linnean Society73:345–390.

Sanmartín I., Ronquist F. (2004) Southern Hemisphere biogeography inferred by event–based models: plant versus animal patterns. Systematic Biology53: 216–243.

Schliep KP. 2010. phangorn: phylogenetic analysis in R. Bioinformatics 27:592-593.

Soltis DE, Soltis PS. 2016. Mobilizing and integrating big data in studies of spatial and phylogenetic patterns of biodiversity. Plant Diversity 38:264-270.

Steel MA, Penny D. 1993. Distributions of tree comparison metrics—some new results. Systematic Biology 42:126-141.

Stucki D, Brites D, Jeljeli L, Coscolla M, Liu Q, Trauner A, Fenner L, Rutaihwa L, Borrell S, Luo T, et al. 2016. Mycobacterium tuberculosis lineage 4 comprises globally distributed and geographically restricted sublineages. Nat Genet 48:1535-1543.

Swofford, D.L. (2003) PAUP*.Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4. Sinauer

Associates, Sunderland, Massachusetts, US.

Yan H-F, Zhang C-Y, Anderberg AA, Hao G, Ge X-J, Wiens JJ. 2018. What explains high plant richness in East Asia? Time

and diversification in the tribe Lysimachieae (Primulaceae). New Phytologist 219:436-448.

Yu Y, Harris AJ, Blair C, He X. 2015. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography.

Molecular Phylogenetics and Evolution 87:46-49.