

# A Rough Guide to RASP 3.1

03/15/2014

Yan Yu<sup>1,2</sup>, AJ Harris<sup>3</sup>, Christopher Blair<sup>2,4</sup>, Xingjin He<sup>1,\*</sup>

<sup>1</sup>*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](mailto:sculab@gmail.com)*

<sup>2</sup>*Department of Biology, Duke University, Box 90338, BioSci 130 Science Drive, Durham, NC, 27708, USA*

<sup>3</sup>*Department of Botany, Oklahoma State University, 301 Physical Science, Stillwater, Oklahoma 74078-3013 USA. Email: [aj.harris@okstate.edu](mailto:aj.harris@okstate.edu)*

<sup>4</sup>*Present address: Department of Biological Sciences, New York City College of Technology, The City University of New York, 300 Jay Street, Brooklyn, NY 11201, USA.*

# Contents

A Rough Guide to RASP 3.1.....	1
1. Introduction (back to contents).....	3
1.1 Explanation.....	3
1.2 Citation .....	3
2. Installing (back to contents) .....	4
3. A general tour of RASP(back to contents) .....	5
3.1 Preparation.....	5
3.2Check status .....	6
3.3 Analysis.....	7
4. How to ...(back to contents) .....	19
4.1 How to make Trees data set.....	19
4.2 How to make a Condensed tree.....	20
4.3 How to make Distributions file.....	21
4.4How to combine results .....	22
4.5 How to remove Outgroups.....	22
5. Explanation of messages (back to contents).....	23
6. References(back to contents).....	24

# 1. Introduction [\(back to contents\)](#)

## 1.1 Explanation

We have written RASP (Reconstruct Ancestral State in Phylogenies) software to implement 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*. DOI:10.1016/j.ympev.2015.03.008

### 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 under mixed models. *Bioinformatics* 19:1572–1574.

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

### **For windows users:**

Supported Operating Systems: Windows 2000 and above(Recommend Win 7)

### **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 XP](#) and should be installed prior to using RASP.

### **For Mac users:**

Supported Operating Systems: 10.6.x and above (Recommend 10.10.x).

### **System Requirement:**

This version of RASP is developed for use on Mac OS X versions 10.6 and above on computers with Intel processors only. 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

Visiting “\” 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 Ware Fusion (commercial) or VirtualBox (free).

### **For Linux users:**

There are very few people using the Linux version of RASP, so we will not update the old version of RASP for Linux. RASP has been not officially supported on the Linux platform since 2012. RASP could be run successfully on Linux using WINE.

### **IMPORTANT NOTE FOR ALL PLATFORMS:**

**Please make sure that the decimal symbol of your system is ‘.’**

# 3. A general tour of RASP [\(back to contents\)](#)

## 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. Distributions file (not required).

(A sample distributions file: “[RASP folder]\Sample\Rubiaceae\RASP\distribution.csv”)

-[How to]: [How to input distributions in RASP.](#)

-[How to]: [How to make a 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] 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], 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 From Trees Dataset] to build one.

**Note1:** The consensus tree computed by RASP do not contain branch length, so it could **NOT** be used for Lagrange, Bayes-Lagrange and BayArea analysis (could be used in S-DIVA and BBM). We strongly recommend user to use their won condensed tree with **branch length**.

**Note2:** 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).

3. If you have a [distributions file](#): Open [File > Load Distribution], 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 distributions in the entry fields in the Distribution column in RASP.

### 3.2 Check status

Tree Option

Binary trees:

Amount of trees:

Discard trees:

Random trees

Current Condensed Tree:

**PLEASE CHECK THE STATUS BEFORE ANALYSIS**

**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** and **Bayes-Lagrange** 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 data setto run the **Statistical Dispersal-Vicariance Analysis** and **Bayes-Lagrange**. 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 Analysis

#### 3.3.1. Statistical Dispersal-Vicariance Analysis

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

Choose [Analysis > Statistical DispersalVicariance Analysis (S-DIVA)] to use default options to run Analysis.

The screenshot shows the S-DIVA software interface. On the left is a range matrix with columns A, B, C, D, E and rows A, B, C, D, E. The matrix contains checkmarks in the following cells: (A,B), (A,C), (A,D), (A,E), (B,C), (B,D), (B,E), (C,D), (C,E), (D,E). Below the matrix is an 'Optimize' section with the following options: 'Max areas at each node' (set to 2), 'Allow Extinction (Slow)' (unchecked), 'Allow Reconstruction (Slow)' (checked), 'Max Reconstructions' (set to 100), 'Random Reconstructions' (checked), and 'Max Reconstructions for final tree:' (set to 1000). On the right is a list of ranges: AB, AC, AD, AE, BC, BD, BE, CD, CE, DE. Above the list are 'Include' and 'Exclude' buttons with arrows. Below the list are 'OK' and 'Cancel' buttons.

#### [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  $i/D_i$  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.

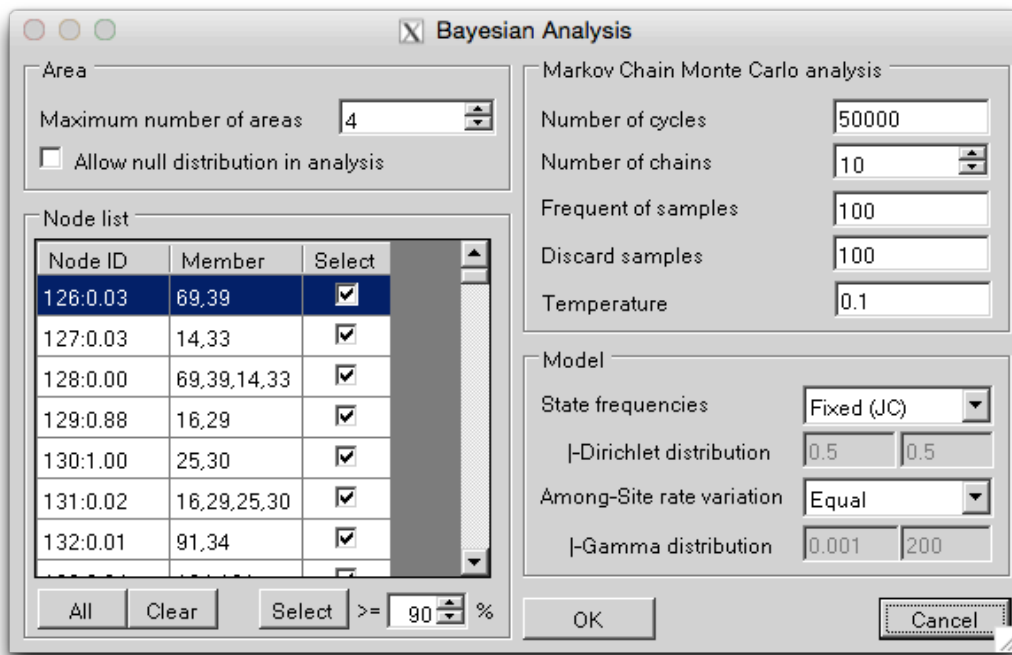
**Random Reconstructions:** Save random reconstructions. We strongly recommend kept this option checked or it will cost a very long time and a large disk space to do the analysis.

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

### 3.3.2. Bayesian Binary MCMC (BBM) Method For Ancestral State

**Note:** BBM Analysis will deal with CONDENSED TREE ONLY

Choose [Analysis > 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]

**Note:** The MCMC analysis is built from source code of MrBayes 3.1.2. We recommend reviewing these options in the Command Reference for MrBayes (Ronquist & Huelsenbeck, 2005)

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:

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

BBM does not consider the branch length and support value of the tree, and it accepts trees with polytomies directly giving the researcher more flexibility in analysis. Researchers also have the ability to collapse branches 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]**

**Note:** There are four models in Bayesian Binary MCMC Analysis: JC, JC+G, F81, F81+G

**State frequencies:** Character state frequencies.

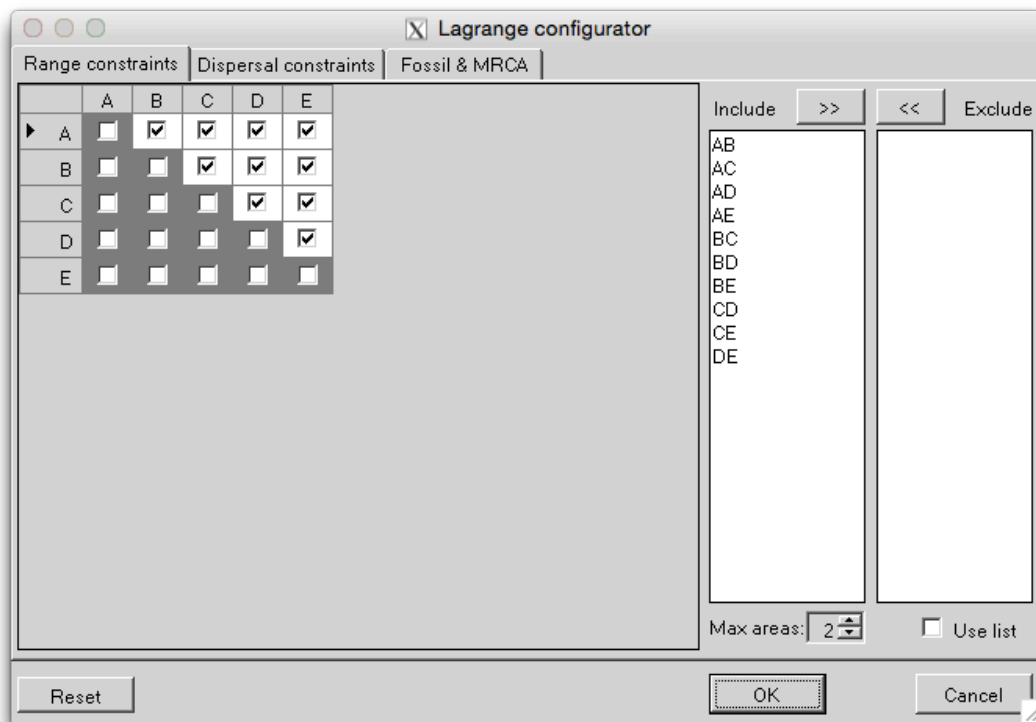
**Among-site rate variation:** This parameter specifies the prior for the gamma shape parameter for among-site rate variation.

### 3.3.3. Lagrange (Dispersal-Extinction-Cladogenesis, DEC model)

**Note:** Lagrange Analysis will deal with CONDENSED TREE ONLY

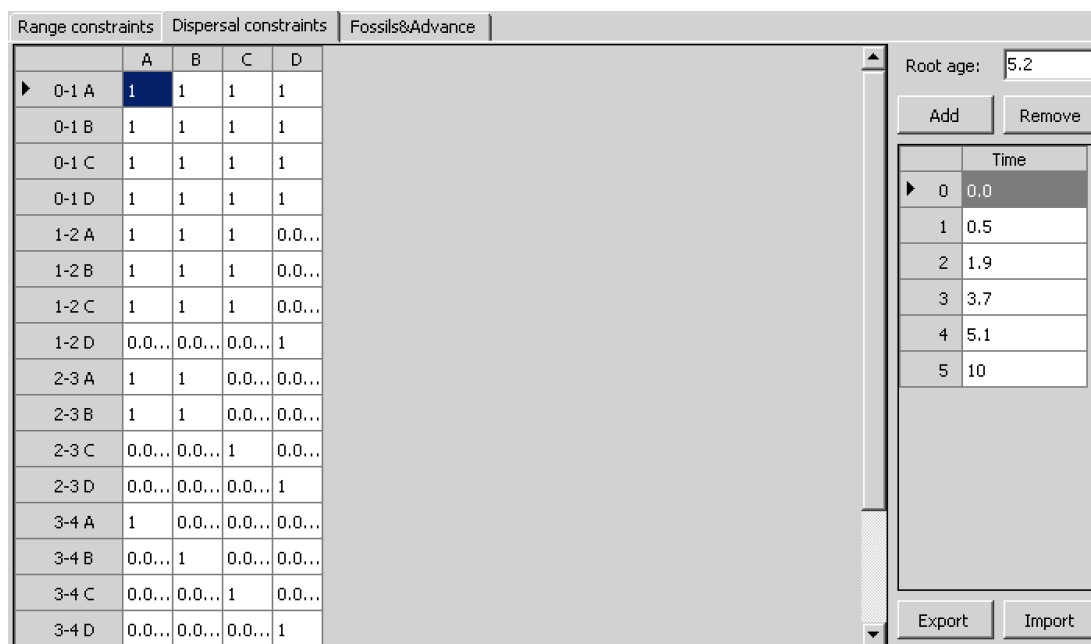
**Note:** For Lagrange and Bayes-Lagrange, we use biogeography of the Hawaiian shrub *Psychotria*, as a sample dataset. You could load the data from “[RASP folder]\Sample\Psychotria”

Choose [Analysis >Lagrange (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

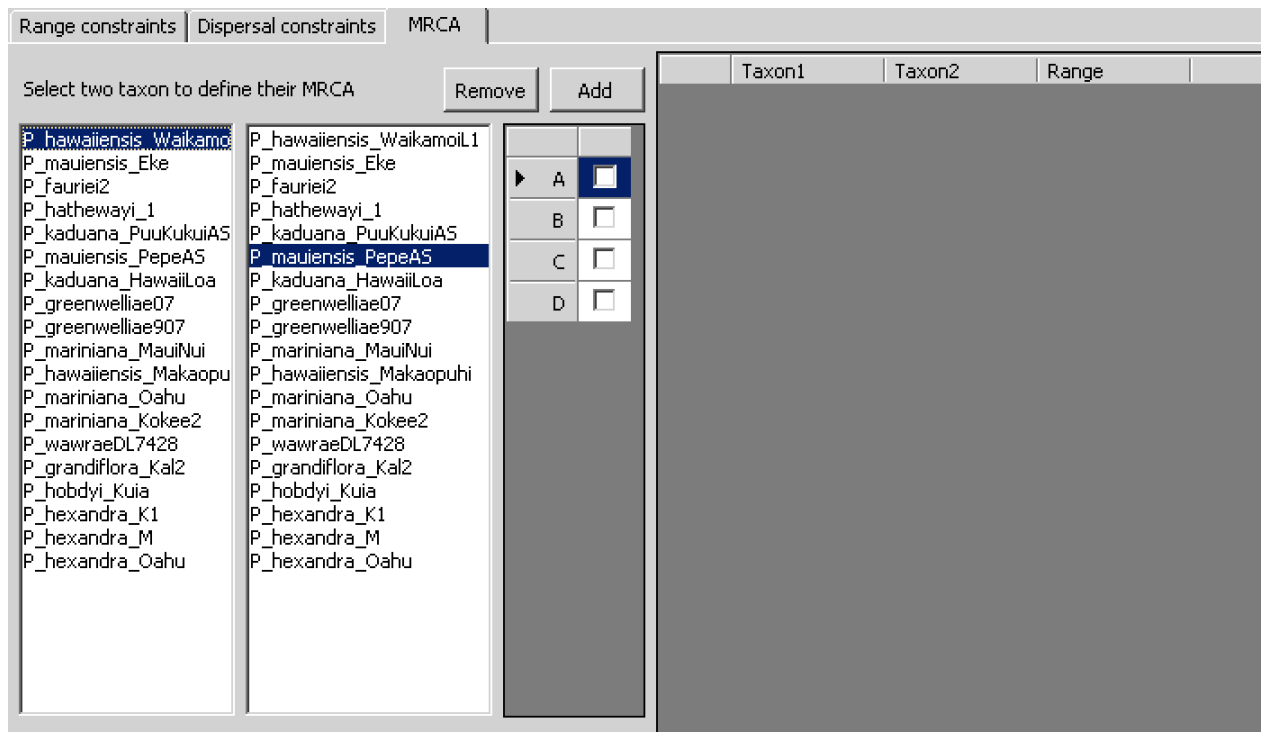
	A	B	C	D
0-1 A	1	1	1	1
0-1 B	1	1	1	1
0-1 C	1	1	1	1
0-1 D	1	1	1	1
1-2 A	1	1	1	0.0...

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 dispersal rates are equal to 1.

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

[Example: Open “[RASP folder]\Sample\Psihotria\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).

### 3.3.4. Bayes-Lagrange (Statistical DEC Model)

Choose [Analysis > Bayes-Lagrange (Statistical DEC, S-DEC)] to use Bayes-Lagrange model

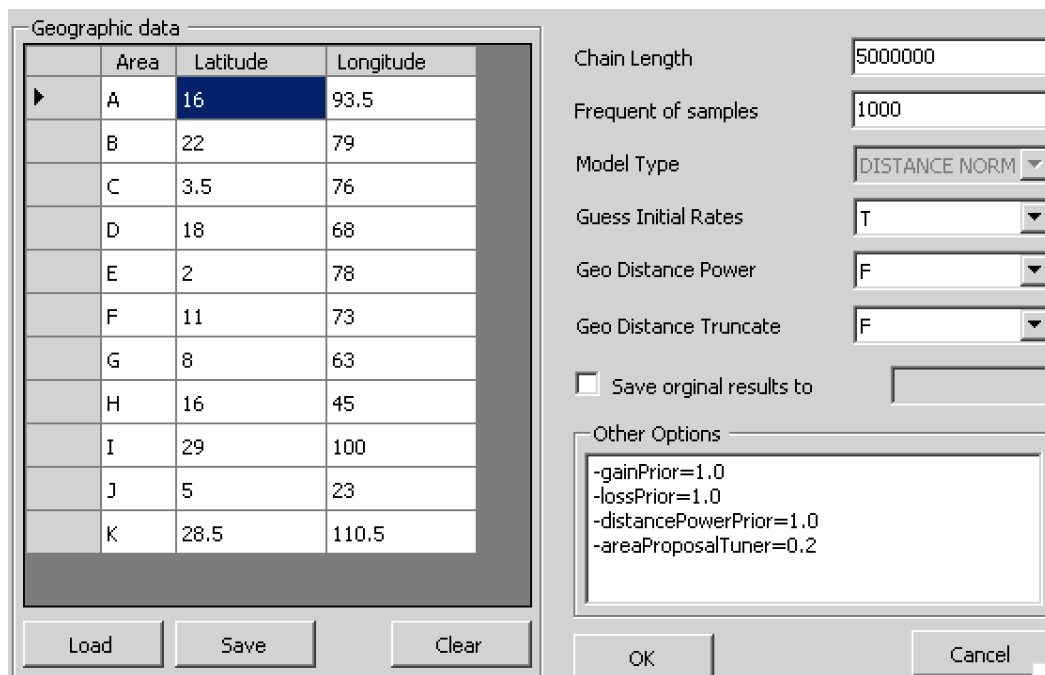
**Note: Bayes-Lagrange Analysis will deal with BOTH CONDENSED TREE AND TREES DATASET**

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.

### 3.3.5. Bayarea (Bayesian inference for discrete areas)

Choose [Analysis > BayArea (Bayesian inference for discrete Areas)] to use Bayes-Lagrange model

**Note:** Bayarea Analysis will deal with CONDENSED TREE ONLY



If you have a Geographicdata file, click [Load] button, navigate to your file and select it. [Example: Open “[RASP folder]\Sample\Rubiaceae\RASP\coordinates.csv” in folder Sample] You can also input and revise the coordinates in the entry fields in the Geographicdata column. The following settings are described in BayArea’s manual (Landis, 2013):

**Chain length:** Number of MCMC cycles.

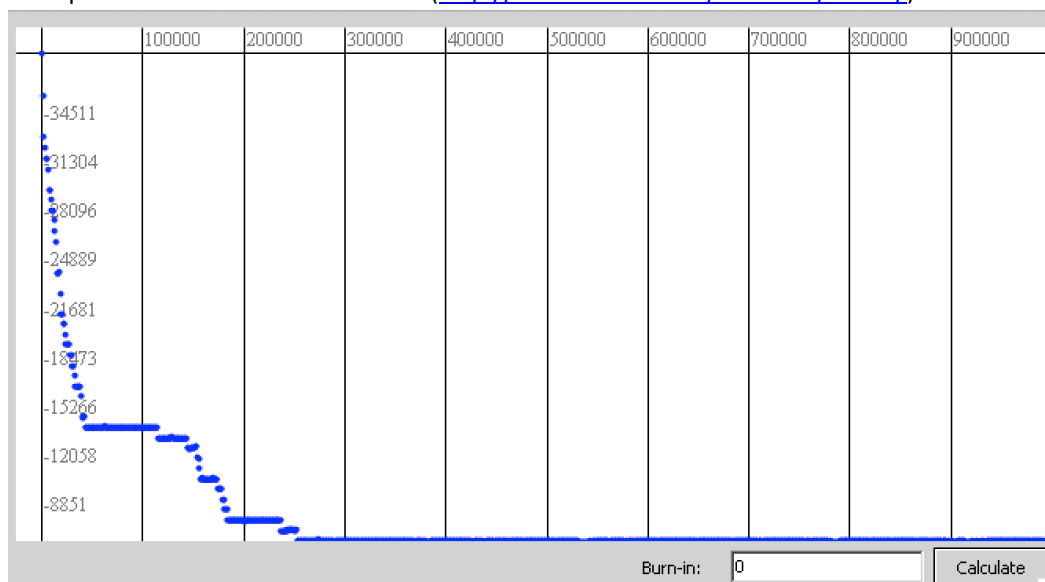
**Frequent of sample:** The MCMC sample frequency to populate area states and probabilities.

**GuessInitialRates:** 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.

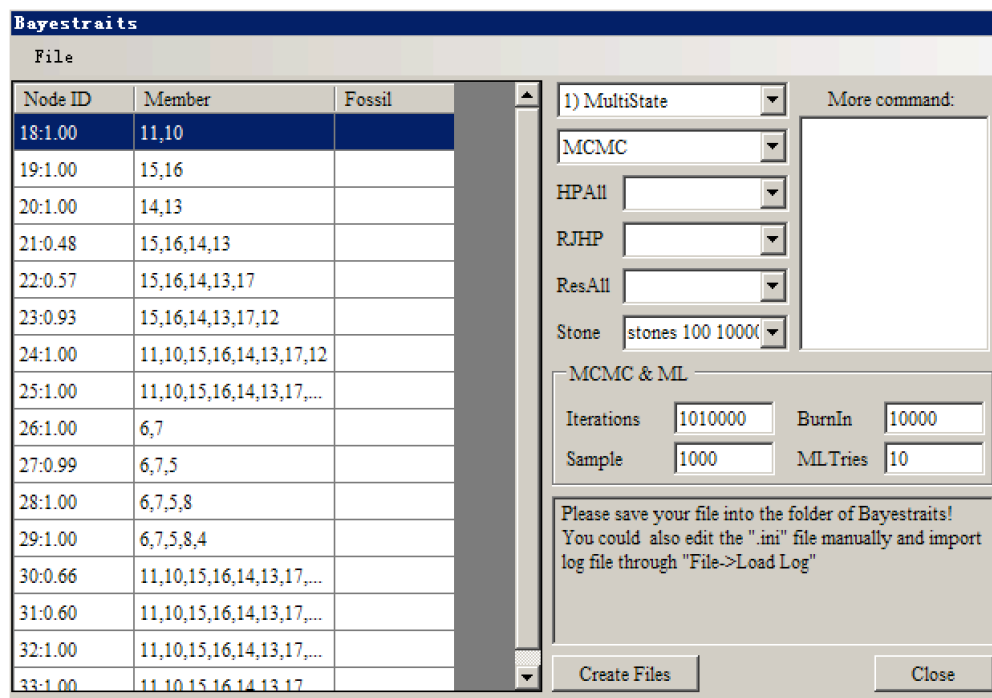
**Geo Distance Truncate:** When True, the pairwise distance calculator approximates extremely improbable dispersal events as having a very small positive rate (effectively zero). This introduces negligible error into the analysis, but greatly speeds up computations when the number of areas is large.

**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/>).

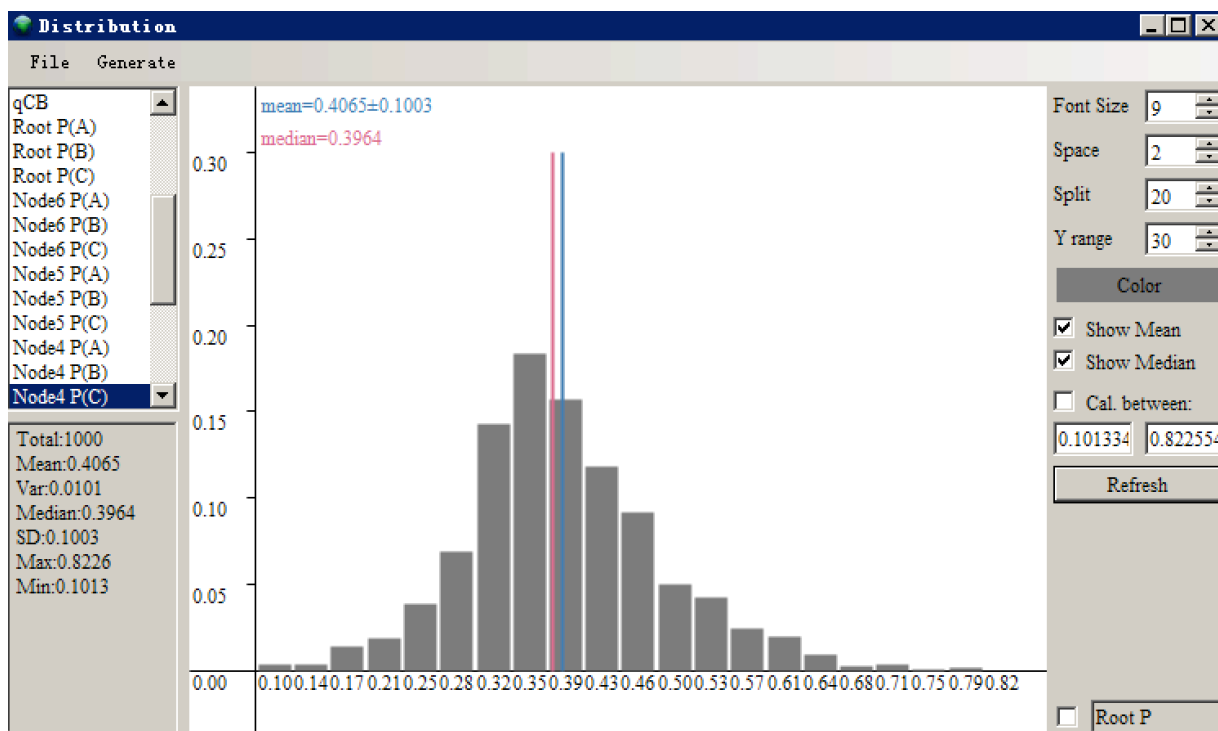


After the MCMC cycle is completed, a “tracer view” will show. Before the calculation of the posterior, the chain must first burn-in. A small Burn-in value will give inaccurate results

### 3.3.6 BayesTraits Command Builder



The BayesTraits Command Builder could build command for BayesTraits. If you want to convert the export of BayesTraits to RASP format, please code the characters with A, B, C etc. and use MultiState Model. After the commands are ready, please click "Create Files" and save your file in the folder of BayesTraits. In the Windows Version of RASP, BayesTraits could run automatically. In the mac version of RASP, you need to run BayesTraits manually, and load the \*.log.txt file from [File->Load Log] in BayesTraits window. You could also analysis the \*.log.txt file from [View->Distribution View] in the main window:



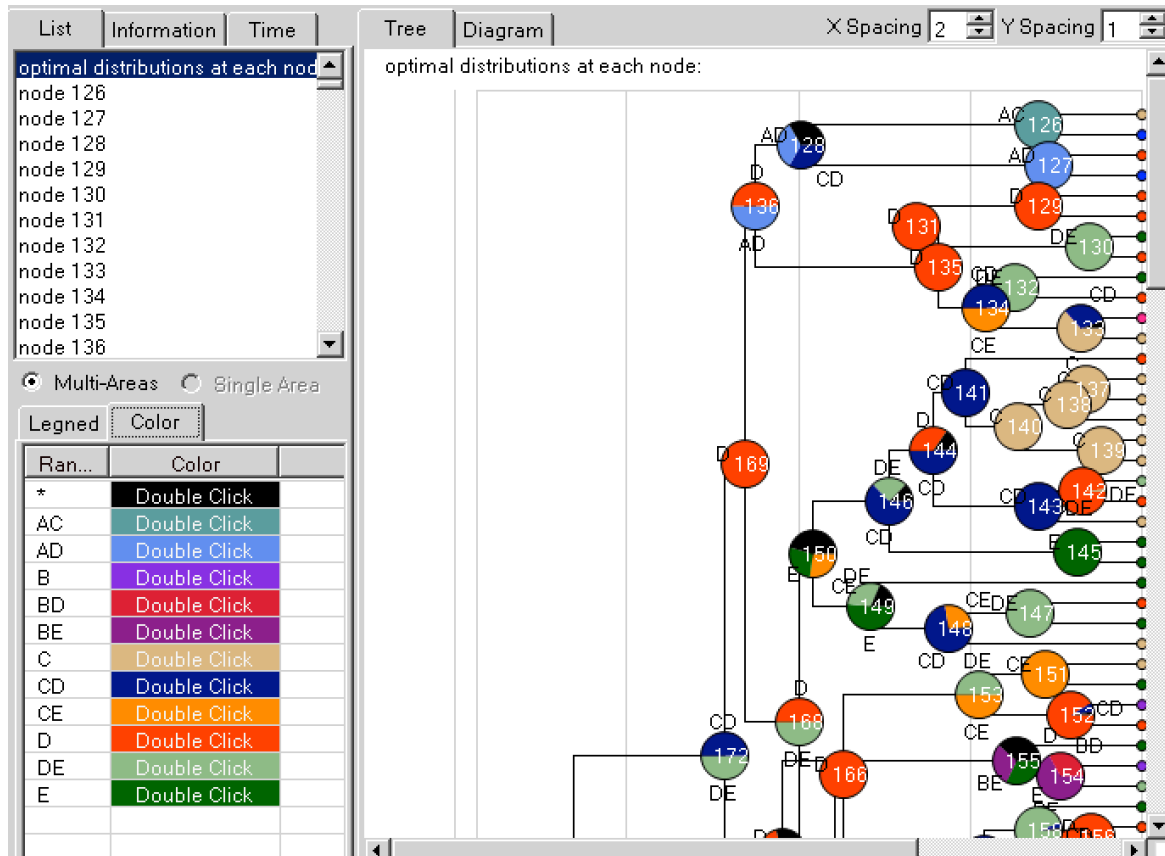
### 3.4 Results

#### 3.4.1. Save and Load Results

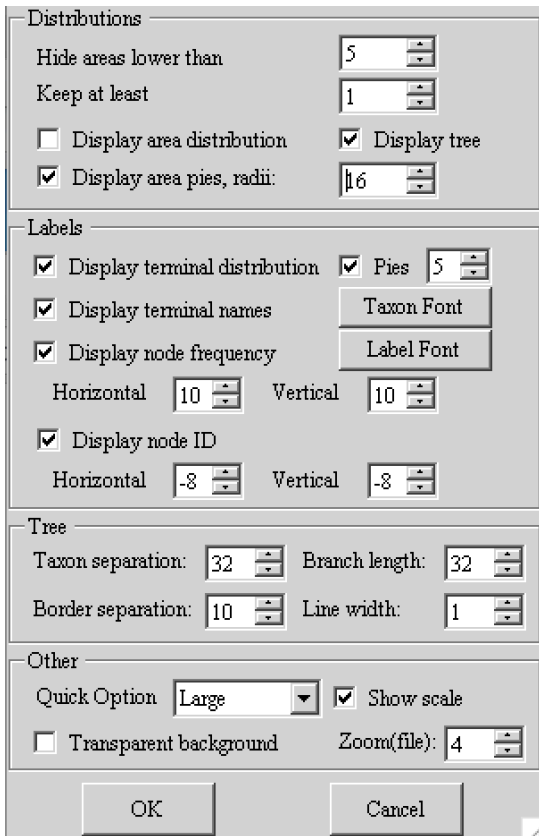
After running an analysis, you can save the last Result from [File > Save > RASP Result]. The result file could be load into RASP from [File > Load Result]. To combine several results, see part 4.4.

#### 3.4.2 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.

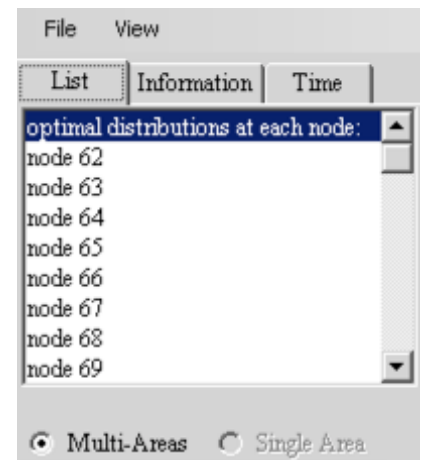
**Border separation:** Set the width of the white space in the tree view window (10-1000). Note that it is possible to make this too small for the width of the tree.

**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 the tree.

**Muti Areas, Single Area:** two different kind of distribution range pie. The Single Area only supports result of BBM and BayArea method. 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  $N_1, N_2 \dots N_j$ , 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:

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

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:

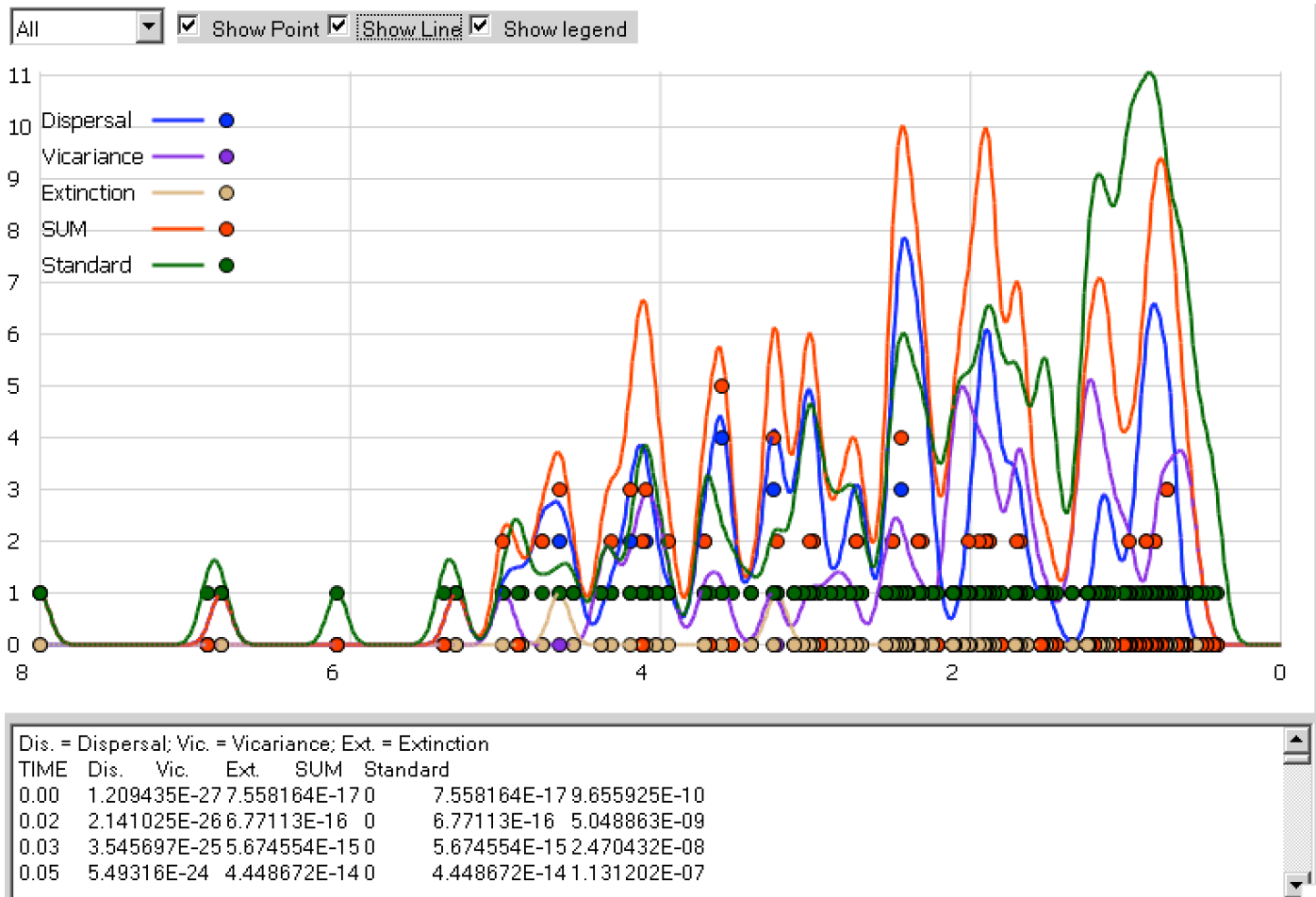
$$D(t) = \sum_{i=1}^n D_i \left( e^{-\frac{(t-T_i)^2}{2}} \right) \quad E(t) = \sum_{i=1}^n E_i \left( e^{-\frac{(t-T_i)^2}{2}} \right)$$

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

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

In these four functions, D(t), E(t), V(t) and M(t) imply 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.

**Note: One could export graphic to vectorial drawing (.svg/wmf file, open with Adobe Illustrator) or picture (.png file).**

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

### 3.4.3Node 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 figure shows 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.

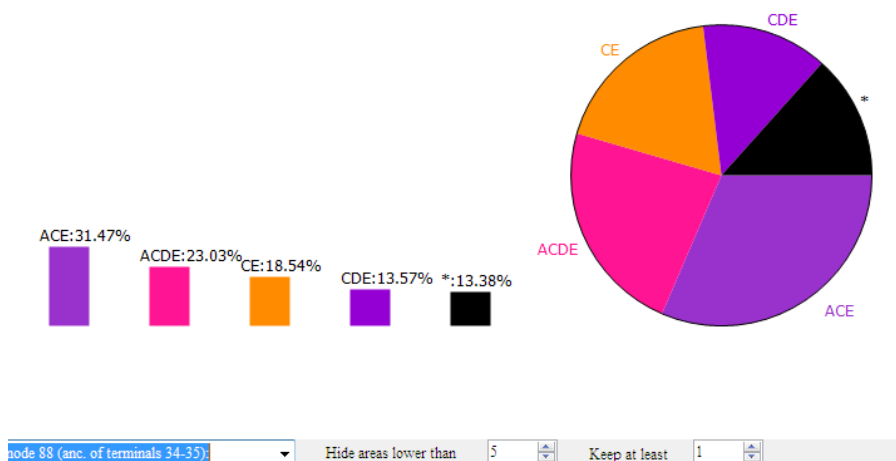
**Note:** Hidden probabilities are lumped together and reported as \* (see far right bar in the figure to the left). In this example, ranges with probabilities less than 5% are hidden.

**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.

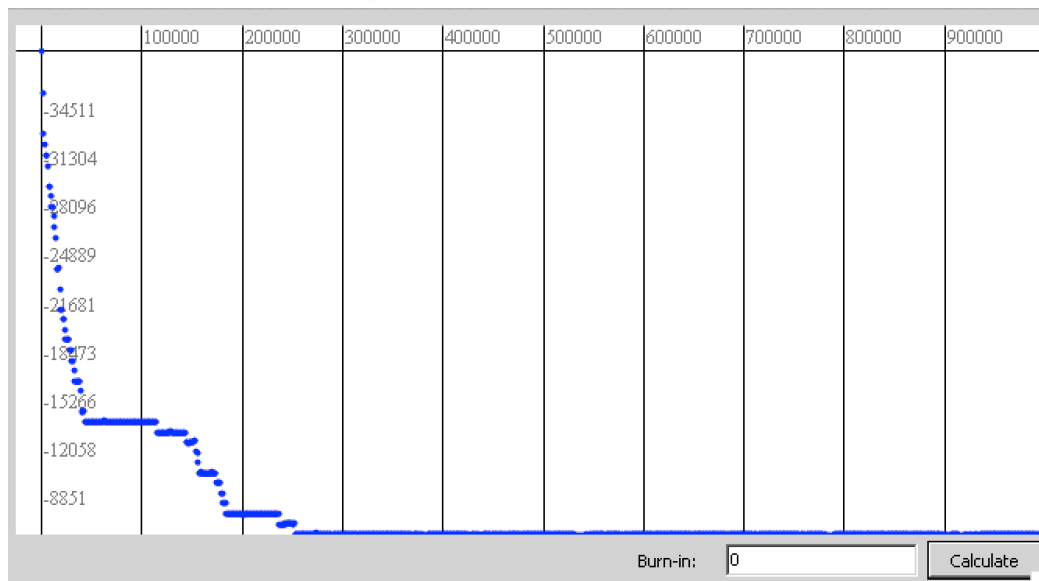
node 88 (anc. of terminals 34-35):



### 3.4.4 Tracer View

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

This can be accessed from [Graphic > Tracer View]



## 4. How to ...[\(back to contents\)](#)

*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 *sample1.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.

**Important: Trees with polytomies could only be used in Bayesian Binary MCMC and MP analysis!**

#### 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.

**Important: Trees with polytomies could only be used in Bayesian Binary MCMC and MP analysis!**

#### 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.

## 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 tree using PAUP. ([Back to Top](#))

- Define outgroups and root 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 Annotator](#) 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=((((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!**

```
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;
```

#### 4.2.4 The 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=((((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:

	A	B	C	D
1	1 Species_01	A		
2	2 Species_02	A		
3	3 Species_03	A		
4	4 Species_04	AB		
5	5 Species_05	B		
6	6 Species_06	B		

```
1, Species_01, A
2, Species_02, A
3, Species_03, A
4, Species_04, AB
5, Species_05, B
6, Species_06, B
7, Species_07, B
8, Species_08, BC
9, Species_09, A
```

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 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 **Distribution** 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 Distribution**.

ID	Taxon	Distribution	Outgroup
1	Strychnos_nuxv...	K	<input type="checkbox"/>
2	Geniostoma_rup...	K	<input type="checkbox"/>
3	Thevetia_peruvi...	ABCDEFGHJ	<input type="checkbox"/>
4	Gelsemium_sem...		<input type="checkbox"/>
5	Phyllis_nobla		<input type="checkbox"/>
6	Paederia_foetida		<input type="checkbox"/>
7	Psvchotria_kirkii		<input type="checkbox"/>

Option

Binary trees (S):

Amount of trees (B):

Discard trees (S):

Random trees (S)

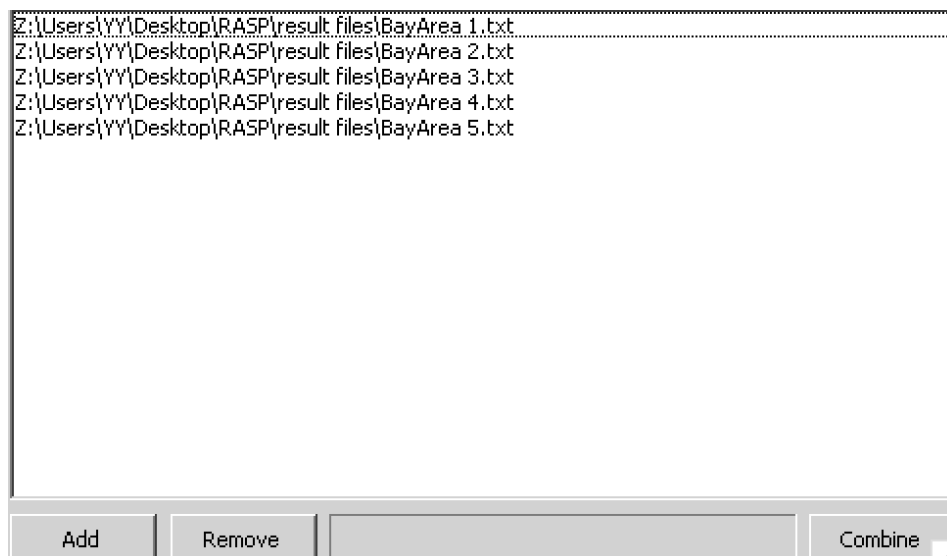
Use Condensed Tree

((((1:44,2:44)1.00:21.4,3:65.4)0.98:3.4,4:68.8

## 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 could close the current data [**File > Close Current Data**] and reload the new data into RASP.

File Graphic Analysis Tools Help				
	ID	Taxon	Distribution	Outgroup
	1	P_hawaiiensis_Wa..	C	<input type="checkbox"/>
	2	P_mauiensis_Eke	C	<input checked="" type="checkbox"/>
	3	P_fauriei2	B	<input checked="" type="checkbox"/>
	4	P_hathewayi_1	B	<input checked="" type="checkbox"/>
▶	5	P_kaduana_PuuK...	C	<input checked="" type="checkbox"/>
	6	P_mauiensis_Pepe..	C	<input type="checkbox"/>
	7	P_kaduana_Hawaii	B	<input type="checkbox"/>

## 5. Explanation of messages [\(back to contents\)](#)

### 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

$$F(x_n)_i = \frac{i}{D_i} .$$

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 [\(back to contents\)](#)

- 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 America* **106** (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.
- 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: Biology* **359**: 1633–1644.
- Drummond, A.J. and Rambaut, A. (2006) BEAST v1.4. <http://beast.bio.ed.ac.uk/>.
- Felsenstein, J. (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.
- 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 Evolution* **47**: 349–368.
- Huelsenbeck J.P., Ronquist F. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.
- 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
- Lamm K.S., Redelings B.D. (2009) Reconstructing ancestral ranges in historical biogeography: properties and prospects. *Journal of Systematics and Evolution*, **47**, 369–382.
- Nylander J.A.A., Olsson U., Alström P, Sanmartín I. (2008) Accounting for phylogenetic uncertainty in biogeography: a Bayesian approach to Dispersal–vicariance Analysis of the thrushes (Aves: Turdus). *Systematic Biology*, **57**, 257–268.
- Ronquist, F. (2001) DIVA version 1.2. Computer program for MacOS and Win32. Evolutionary Biology Centre, Uppsala University. Available at <http://www.abc.uu.se/systzoo/research/diva/diva.html>.



Ronquist, F. (1997) Dispersal-vicariance analysis: A new 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.abc.uu.se/systzoo/research/diva/diva.html>.

Ronquist F, Huelsenbeck JP (2003) MrBayes3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*

19:1572–1574.

Sanmartín I, Enghoff H., Ronquist F. (2001) Patterns of animal dispersal, vicariance and diversification in the

Holarctic. *Biological Journal of the Linnean Society* **73**:345–390.

Sanmartín I., Ronquist F. (2004) Southern Hemisphere biogeography inferred by event-based models: plant versus animal

patterns. *Systematic Biology* **53**: 216–243.

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

Associates, Sunderland, Massachusetts, US.