RASER | Version 2.0

The RAte Shift EstimatoR (RASER) is a Bayesian method for testing and detecting site-specific evolutionary rate shifts. Given a multiple sequence alignment (MSA) and a phylogenetic tree, the program determines whether or not site-specific rate shifts characterize the evolution of a protein, and if so, points to the specific sites and lineages in which these shifts have most likely occurred.

Manual

Contents

New features

This new release includes two additional features:

1. Stochastic mapping of mutations (Nielsen R. 2002. *Syst. Biol.* 51(5):729-739) is implemented to calculate more accurately the probability that a rate-shift occurred at a specific branch.
2. Parameters text file is now used as input instead of reading the program parameters (or options) from the command line.

Download and Installation

Windows and Linux executable files and source code (C++) are available at

http://www.tau.ac.il/~penn/raser.html

Compiling RASER

1. In order to unzip and untar the files please type:
   tar -xzvf raser.v2.0.tar.gz
   This will create the following directories:
   libs/phylogeny
   programs/raser
2. In some operating systems, you may use the makefiles to compile the program. If this does not work, skip to item 3.
   Make sure you are in the directory where you unzipped the files, and type:
   cd libs/phylogeny
   In order to run the Makefile, type:
   make
   Now, type:
   cd ../../programs/raser
   to get to the raser directory. Type:
   make
   in order to run the Makefile.
   This will result in an executable file called *raser* which will reside in the *programs/raser* directory.
3. In some systems (such as Unix), the makefiles will not be operable. Thus, follow step 1 and compile directly using g++:
   a. Make sure you are in the directory where you unzipped the files.

b. Type:
mv libs/phylogeny/ programs/raser/*
c. cd to the raser library:
cd programs/raser
d. To compile, type
*g++ -O3 -o raser *.cpp*
This will result in an executable file called raser which will reside in the src/raser directory.

If there are any problems with the compilations (occasionally, with old version of g++) - please email penn@post.tau.ac.il and I'll try to help. To modify the code, or use parts of it for other purposes, permission is requested. Please contact Tal Pupko at talp@post.tau.ac.il. Please note that the use of the RASER program is for academic use only.

Running RASER

In order to infer rate shifts for a certain dataset, we recommend performing the following stages according to the most suitable scenario:

A. To infer rate shifts across the entire phylogenetic tree, without focusing on a specific lineage:
    1. Perform a likelihood ratio test. If this test is in support of rate shifts, proceed to the next two stages.
    2. Infer rate-shifting sites.
    3. Infer rate-shifting lineages for rate-shifting sites.
B. To infer rate shifts in a specific lineage:
    1. Specify the branch.
    2. Perform a likelihood ratio test. If this test is in support of rate shifts, proceed to the next stage.
    3. Infer rate-shifting sites in the specific lineage.

To run the program you must supply a parameters text file. Simply type in the command line:

```
raser parameters_file_name
```

For the above mentioned two scenarios, example parameters files are available at the RASER webpage: http://www.tau.ac.il/~penn/raser.html. See below how to use them.

For more complex options see the `raser.allOptions.params` file, also available at the RASER webpage.

Answer the following question in order to choose the most suitable scenario:

Do you have a suspect branch you want to test?

A. No suspect branch - infer rate shifts across the entire phylogenetic tree

1. Performing a likelihood ratio test

In order to perform a likelihood ratio test, the program must be run twice: once with a rate-shift enabling model and once with a null model (which does not enable rate shifts). Then the likelihood values of the two runs must be compared (twice the difference of the likelihood values follows a chi-square distribution).
Thus, run the program twice, with the following parameters files (make sure to use the same model name in both runs):

1. For the rate-shift model use the file: `raser.params`.
2. For the null model to run, use the file: `null.params`.

The likelihood of the data given each model will be in the results file.

2. Inferring rate-shifting sites

The results file of the rate-shift model from stage 1 will contain all the information required for determining which sites have experienced rate shifts. The file will look like this:

```
RASER Results File
============================================================
Parameters are:
Log-likelihood: -27691.8
Alpha for rate distribution: 0.298176
Prob (nu=0): 0.605
Alpha for nu distribution: 1.44419
Beta for nu distribution: 1
Prob (nu=infinity): 0.0627223
Rate-Shifting Positions
Displayed on sequence HXB2.B
```

```
================================================================================
POS     AMINO   POSTERIOR PROBABILITY OF RATE-SHIFT (* IF > 0.95)
1       M       0.097
2       E       0.38
3       P       0.96    *
4       V       0.95
5       D       0.11
6       P       0.99    *
```

The results file, as exemplified above, presents the posterior probability that each site has undergone a rate shift. Sites for which this probability is higher than the 0.95 threshold are marked with an asterisk.

3. Inferring rate-shifting lineages

For any site suspect of having undergone rate shifts from stage 2, we would like to know the lineage/s in which the rate shift occurred. This information will be displayed in the nodes_results_file (the name of the file is defined using the _outNodesResFile parameter in the parameters file). A typical file will look like this:

```
pos    node   probAcc      probDec      probShift

3      16     0.0395 0.0002 0.0397

3      1      0.0035 0.0220 0.0255

3      12     0.0136 0.0002 0.0137

6      1      0.0014 0.0088 0.0102

6      108    0.0030 0.0007 0.0037

6      139    0.0026 0.0010 0.0036
```

For each site with posterior probability of rate-shift higher than the cutoff (marked with '*' in the results file of step 2), the three lineages (nodes) with the highest posterior probability of a rate shift at this lineage are presented. Furthermore, for each such lineage the probability of a rate-acceleration and a rate-deceleration are presented (note that if the input tree is unrooted, these terms are only meaningful in relation to one another).

The information presented in this file refers to node numbers along the tree. In order to be able to view which lineage corresponds to which node number, a file is automatically created with the name of the _outTreeFileWithBranchesNames parameter. This file is the tree in Newick format, with the node numbers displayed as bootstrap values. The tree may be viewed with software such as NJplot (Perrière, G. and Gouy, M. 1996, http://pbil.univ-lyon1.fr/software/njplot.html).

B. With a suspect branch - infer rate shifts in a specific lineage

1. Specify the branch

Run RASER with the parameters file: *onlyPrintNodesNames.params*.
The *_outTreeFileWithBranchesNames* parameter defines the file name, which includes the tree in Newick format, with the node numbers displayed as bootstrap values. The tree may be viewed with software such as NJplot (Perrière, G. and Gouy, M. 1996, http://pbil.univ-lyon1.fr/software/njplot.html). You should find the relevant branch in the tree and use it in the next step.

2. Performing a likelihood ratio test

In order to perform a likelihood ratio test, the program must be run twice: once with a rate-shift enabling model and once with a null model (which does not enable rate shifts). Then the likelihood values of the two runs must be compared (twice the difference of the likelihood values follows a chi-square distribution).
Thus, run the program twice, with the following parameters files (make sure to use the same model name in both runs):

1. For the rate-shift model use the file: `raser.stochasticMapping.params`. Specify the branch number from stage 1 in the *_specificNodeForStochasticMapping* parameter. Using this parameters file, you are not only calculating the likelihood of the rate-shift model, but also running a stochastic mapping algorithm (Nielsen R. 2002. *Syst. Biol.* 51(5):729-739) on the specific branch. In essence, this algorithm provides more accurate results in the next stage (stage 3).
2. For the null model to run, use the file: `null.params`.

The likelihood of the data given each model will be in the results file.

3. Infer rate-shifting sites in the specific lineage

This stochastic mapping algorithm results will be displayed in the nodes_results_file (the name of the file is defined using the _outNodesResFile parameter in the parameters file). For each site the probability of a rate-acceleration and a rate-deceleration in the specific branch (defined in step 1) are presented. Note that if the input tree is unrooted, these terms are only meaningful in relation to one another.

More options and instructions

You may use the *raser.allOptions.params* file that includes a list of all the options below.

The basic options are:

	Name	Description	Default	Remarks
Input	_inSeqFile	Input aligned sequence file	Obligatory	Use full path. Formats accepted are: Fasta, Clustal, Phylip, Mase
	_inTreeFile	Input user tree in Newick file.	NJ tree	Use full path
	_inQuerySeq	Name of query sequence	1st in the sequences file	
Output	_outResFile	Results output file	Obligatory	Use full path
	_outNodesResFile	Linages results output file	Obligatory	Use full path
	_logFile	Log file name		Use full path
	_outTreeFile	Output tree file (with optimized branch lengths)		Use full path

The more complex options are:

	Name	Description	Default	Remarks
Input	_modelName	{jtt (JTT), rev (REV - for mitochondrial genomes), day (DAY), wag (WAG), cprev (cpREV for	jtt	

		chloroplasts genomes), HIVb, HIVw, aajc (JC amino acids), nucjc (JC nucleotides)}		
	_numOfCategoriesForRateDistr	Number of categories for rate discrete distribution	4	Integer
	_numOfCategoriesForNuDistr	Number of categories for nu gamma discrete distribution	5	Integer. Use 4 to make the run faster, although slightly less accurate.
Output	_outTreeFileWithBranchesNames	The tree in Newick format, with the node numbers displayed as bootstrap values	raser.tree.namesBS.ph	Use full path.
	_verboseLevel	Verbose level for log file	5	Integer
Running mode	_onlyPrintNodeIDs	No likelihood calculations	0=false	0 or 1
	_useNullModel	Fix prob(nu=0)+prob(nu=inv)=1	0=false	0 or 1
	_specificNodeForStochasticMapping	Run stochastic mapping algorithm on the specified node	-1=do not run stochastic mapping algorithm	Node number or -1

	_numBranchesToPrint	Number of lineages with the highest posterior probability of a rate shift that are printed for each site with posterior probability higher than the cutoff	3	Integer. Not available for stochastic mapping.
Optimization	_manyStartPointsOpt	Use many starting points for optimization	1=true	0 or 1
	_bblOpt	Perform branch length optimization	1=true	0 or 1
	_numBBLiterations	Number of branch length optimization iterations	5	Integer
	_epsilonLikelihood	Epsilon for likelihood optimization	0.1	The smaller the value, the higher the precision
Cutoffs	_posteriorCutoffSites	Lineages are inferred only for sites with posterior probability higher than the cutoff	0.95	Not available for stochastic mapping
	_posteriorCutoffNodes	Nodes with posterior probability higher than the cutoffs are marked with asterisks	0.5	

Initialize parameters values	_initAlphaNuDistr	Initial probability of alpha of nu distribution	1.0	
	_initBetaNuDistr	Initial probability of beta of nu distribution	1.0	
	_initAlphaRateDistr	Initial probability of alpha of rate distribution	1.0	
	_initProbNu0	Initial probability of nu=0	0.5	
	_initInfinityProb	Initial probability of nu=infinity	0.25	
Fixing parameters	_fixedAlphaNuDistr	Fix alpha nu distribution	0=false	0 or 1
	_fixedBetaNuDistr	Fix beta nu distribution	0=false	0 or 1
	_fixedAlphaRateDistr	Fix alpha rate distribution	0=false	0 or 1
	_fixedProbNu0	Fix probability of nu=0	0=false	0 or 1
	_fixedInfinityProb	Fix probability of nu=infinity	0=false	0 or 1