RAxML-Workbench Help

1 Main Menu Bar

New Submission

Switch to the Submission-section to use RAxML or other tools provided.

Open Job

Open an older RAxML Job directory (RAxML_info file has to be present) and open it within the GUI for further analysis.

Open file in Treeviewer

Display a treefile in PhyloXML format in the Treeviewer.

Preferences

Enable Uclust:

> Download RC Edgar's *uclust*(usearch) from www.drive5.com/usearch/. Select the uclust executable binary with the given file browser. You can also specify the identity level for the clusters in the field "Cluster Identity". Save your changes with the "Save" Button. If your configuration was valid, then the "Cluster Reads" Option within the EPA-Submission forms should be enabled.

Attention! It might happen that programs change their input specifications with new versions. We try to keep track of this, but if you encounter problems feel free to contact us (Please send us also the version of the program you used).

2 The Submission-section

In this Section of the RAxML Workbench, you can run RAxML or other tools.

2.1 RAxML-EPA - Single Gene Alignment

The RAxML-EPA-Pipeline for Single Gene Alignment-files. The first 4 parameters must be entered to let the EPA run:

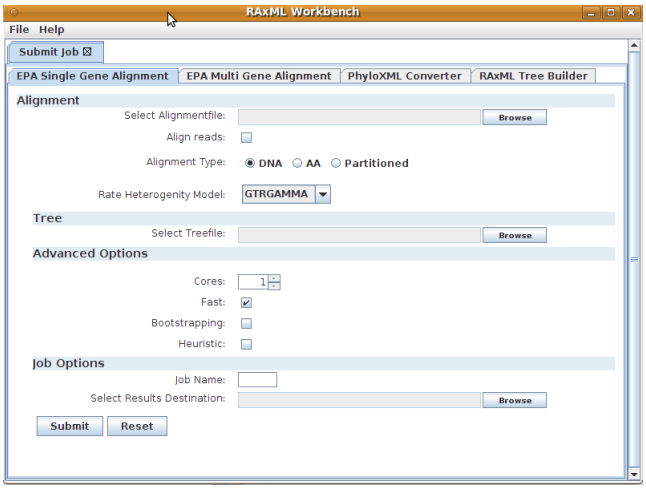

Figure 1: The Submission-section

Alignmentfile

Select a reference alignment with aligned query sequences in FASTA or PHYLIP format. If the query sequences are not aligned to the reference alignment, you can check the *Upload unaligned reads* button below and select the query reads separately. The program will automatically rename taxa if they are not unique within the alignment due to PHYLIP format conversion tools that restrict taxa names to 10 characters only.

Treefile

Select an unrooted, strictly bifurcating reference tree in Newick format that contains the reference sequences. The tree does not need to contain branch lengths because they will be automatically estimated by the RAxML EPA algorithm.

Results Destination

Select the parent folder where the RAxML Workbench is going to save your results.

Jobname

Enter the name of the folder where RAxML is going to save your results.

There are more optional parameters available:

Align Reads

Check this box if you want the program to automatically align your reads to the full-length reference sequence alignment using *hmmalign*. In the case that you have a huge number of reads you may then also chose the option *Cluster reads* (if enabled, see Preferences). The aligning is then only performed for consensus reads that are built using RC Edgar's *uclust* program.

Cores

If you have more than one core available you can increment this number to perform the parallel version of RAxML on your local machine.

Alignment Type

Select the type of data in your alignment, this can be either DNA or Protein data or any combination of DNA, Protein, Secondary Structure, Multi-State, or Binary data partitions as specified by a standard RAxML partitioned model file (passed vie the -q option).

Bootstrapping

Check this button if you want to infer query placement uncertainty values via standard phylogenetic Bootstrapping.

Heuristic

Check this box if you want to use the fast placement heuristics (recommended for large datasets).

2.2 RAxML-EPA - Multi Gene Alignment

The RAxML-EPA-Pipeline for Multi Gene Alignment-files. The first 6 parameters must be entered to let the EPA run:

Alignmentfile

Select a reference alignment with aligned query sequences in FASTA or PHYLIP format. If the query sequences are not aligned to the reference alignment, you can check the *Upload unaligned reads* button below and select the query reads separately. The program will automatically rename taxa if they are not unique within the alignment due to PHYLIP format conversion tools that restrict taxa names to 10 characters only.

Treefile

Select an unrooted, strictly bifurcating reference tree in Newick format that contains the reference sequences. The tree does not need to contain branch lengths because they will be automatically estimated by the RAxML EPA algorithm.

Multi Gene Partitionfile

Select a partitionfile that tells RAxML how the alignment is separated into the different genes.

Readsfile

Select file containing the reads in FASTA format that have to be aligned. The program will assign the reads to the most fitting gene based on the Smith-Waterman algorithm. The reads are then aligned with *hmmalign*. In the case that you have a huge number of reads you may then also chose the option *Cluster reads* (if enabled, see Preferences). The aligning is then only performed for consensus reads that are built using RC Edgar's *uclust* program.

Results Destination

Select the parent folder where the RAxML Workbench is going to save your results.

Jobname

Enter the name of the folder where RAxML is going to save your results.

There are more optional parameters available:

Cores

If you have more than one core available you can increment this number to perform the parallel version of RAxML on your local machine.

Bootstrapping

Check this button if you want to infer query placement uncertainty values via standard phylogenetic Bootstrapping.

Heuristic

Check this box if you want to use the fast placement heuristics (recommended for large datasets).

2.3 Generate PhyloXML files

This tool can generate PhyloXML formated files from your (older) RAxML results, that can be displayed in the treeviewer. The program needs three input parameters to run.

RAxML Classificationweights

Select a RAxML_classificationLikelihoodWeights.-File, a RAxML_classification.-File in case you performed RAxML with bootstrapping or another file that has also 4 columns where the first column is the sequence-name, the second the branch-name, the third the likelihood weight and the fourth the sum of the likelihood weights listed above from the same sequence.

Reference Tree

Select the unrooted, strictly bifurcating reference tree in Newick format on which the RAxML-EPA algorithm performed the placements listed in the corresponding classificationfile selected above.

Destination path

Simply specify where the resulting PhyloXML file should be saved.

2.4 RAxML Tree Builder

RAxML (Randomized Axelerated Maximum Likelihood) is a program for sequential and parallel Maximum Likelihood based inference of large phylogenetic trees. The first 3 parameters must be entered to let RAxML run:

Alignmentfile

Select a reference alignment with aligned query sequences in FASTA or PHYLIP format. If the query sequences are not aligned to the reference alignment, you can check the *Upload unaligned reads* button below and select the query reads separately. The program will automatically rename taxa if they are not unique within the alignment due to PHYLIP format conversion tools that restrict taxa names to 10 characters only.

Results Destination

Select the parent folder where the RAxML Workbench is going to save your results.

Jobname

Enter the name of the folder where RAxML is going to save your results.

There are more optional parameters available:

Cores

If you have more than one core available you can increment this number to perform the parallel version of RAxML on your local machine.

Parsimony Random Seed

Specify a random number seed for the parsimony inference.

Alignment Type

Select the type of data in your alignment, this can be either DNA or Protein data or any combination of DNA, Protein, Secondary Structure, Multi-State, or Binary data partitions as specified by a standard RAxML partitioned model file (passed vie the -q option).

Bootstrapping

Check this button if you want RAxML to perform a bootstrap analysis.

3 The Results-section

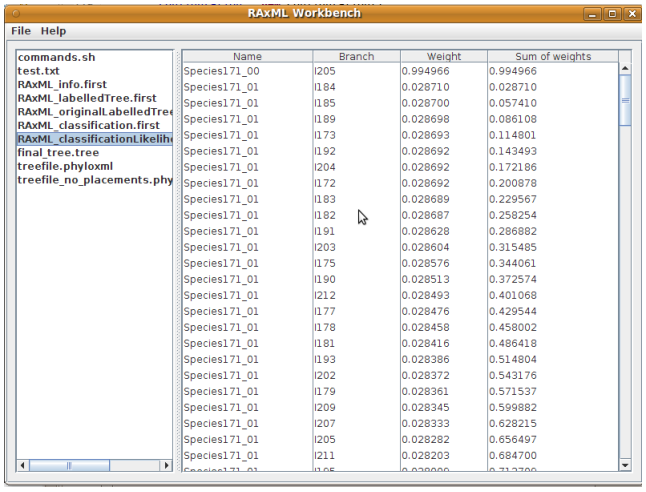

Figure 2: The Results-section

On the left-hand side, all available files within the selected RAxML job-folder are displayed. The content of the files is shown in the right panel when clicking on it. Classificationfiles are displayed as tables and PhyloXML formated files are directly displayed in the treeviewer. Every other file is displayed as plain text.

4 The Treeviewer

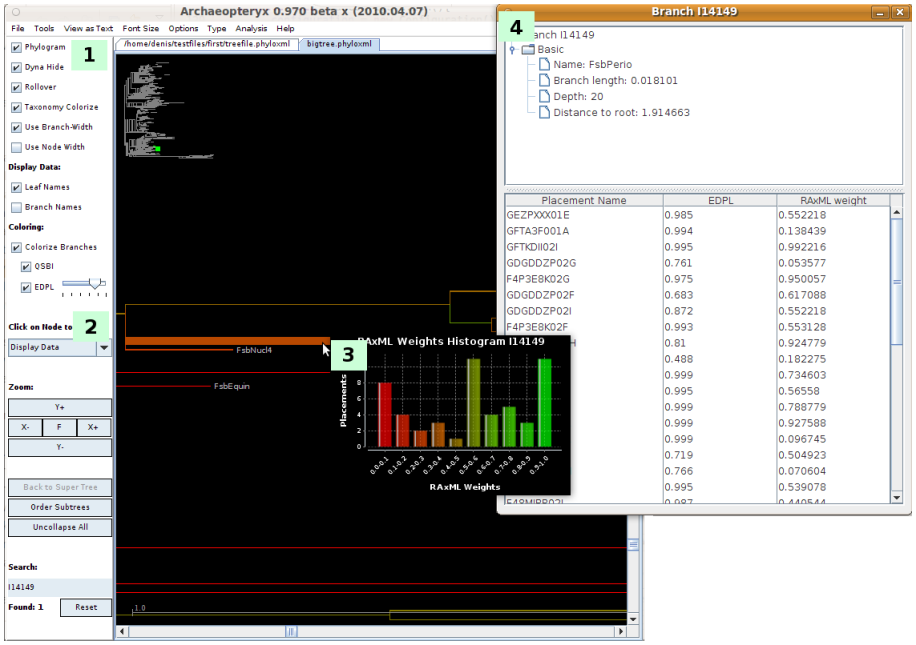

Figure 3: The Archaeopteryx Treeviewer

We provide a brief overview of some relevant Archaeopteryx Treeviewer features below. For a more detailed description, please refer to the Archaeopteryx project web-site http://www.phylosoft.org/archaeopteryx
Attention! The features listed below are not part of the original Archaeopteryx treeviewer and were added subsequently:

- Colorize Branches with different scorings
- Mouse rollover histograms
- Branchdata pop-ups
- Branch labels
- Node width

If you have questions, proposals or other concerns regarding these points, feel free to contact us.

4.1 Left Checkboxes(1)

Phylogram

The branch lengths of the tree correspond to the branch length given within the tree.

Dyna Hide

Dynamically hide labels if not enough space is available.

Rollover

Activates the mouse over function for the tree.

Colorize Branches

Assign colors to branches based on diversity measures. The coloring represents the mean value of the confidences off all placements on that branch. The confidences correspond to the diversity scorings selected beneath.

- *QSBI:* - A prototype diversity measure that scores for every branch, if the placements distribution in the left and right subtree of a query sequence is better than expected. The mean of all query sequences is taken as score. Green means that the placement distributions in the left and right subtree is significantly different from what has been expected.
- *EDPL:* - Measures the spreading of the query sequence within the tree. A brighter green means a very high mean confidence on that branch. A bright red means the opposite. The slider allows a manual adjustment of the cutoff.

All measurements can be combined.

Use Branch-Width

Display the fraction of reads that have been placed into a branch by means of a relative branch width.

Use Node-Width

Display the fraction of reads that have been placed in the underlying subtree by means of a relative node width.

Leaf Names

Display the taxa names on the leaves of the tree.

Branch Names

Display the branch labels.

4.2 Tree-manipulation(2)

In the drop down menu below "Click on Node to:" you can switch between actions that are performed when you click on a node or a branch in the visualized tree respectively.

- Display Node Data
- Collapse/Uncollapse
- Root/Reroot (only on branches)
- Sub/Super Tree
- Swap Descendants
- Colorize subtree
- Delete Subtree/Node

4.3 Top main panel

Type:

Select your preferred tree layout between the following:

- Rectangular
- Euro Type
- Rounded
- Triangular
- Unrooted

View as Text:

View the represented tree in different file formats like e.g. phyloXML

4.4 The interactive tree visualization

Branch placement histogram(3)

A mouse roll over on a branch that has placements on it displays a histogram of RAxML weights. These weights represent either likelihood weights or bootstrap supports.

Branch Data View(4)

By left clicking on a branch, a pop up window is displayed showing a more detailed view of the placements present on the corresponding branch. In the bottom part of the window, a table with the read names, the EDPL scores and the RAxML weights (likelihood or bootstrap support) are displayed.