raxmlGUI 2.0 beta: a graphical interface and toolkit for phylogenetic analyses using RAxML

Daniel Edler [1,2,3], Johannes Klein[2], Alexandre Antonelli[2,3,4,*], Daniele Silvestro[2,3,5,*]

1. Integrated Science Lab, Department of Physics, Umeå University, Sweden.

2. Gothenburg Global Biodiversity Centre, Box 461, SE-405 30 Gothenburg, Sweden.

3. Department of Biological and Environmental Sciences, University of Gothenburg, Carl Skottsbergs gata 22B, Gothenburg 41319, Sweden.

4. Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AE, United Kingdom

5. Department of Computational Biology, University of Lausanne, Quartier Sorge, 1015 Lausanne, Switzerland.

[*] co-last authors.

Correspondence to:

D. Edler <daniel.edler@umu.se>

D. Silvestro <daniele.silvestro@bioenv.gu.se>

Abstract

RaxmlGUI is a graphical user interface to RAxML, one of the most popular and widely used software for phylogenetic inference using maximum likelihood. Here we present raxmlGUI 2.0-beta, a complete rewrite of the GUI, which replaces raxmlGUI and seamlessly integrates RAxML binaries for all major operating systems providing an intuitive graphical front-end to set up and run phylogenetic analyses. Our program offers automated pipelines for analyses that require multiple successive calls of RAxML and built-in functions to concatenate alignment files while automatically specifying the appropriate partition settings. While the program presented here is a beta version, the most important functions and analyses are already implemented and functional and we encourage users to send us any feedback they may have. RaxmlGUI facilitates phylogenetic analyses by coupling an intuitive interface with the unmatched performance of RAxML.

Introduction

Phylogenetic inference is a keystone in evolutionary biology research. It provides the foundations for tackling a wide range of questions, from population dynamics to taxonomy of higher taxa. RAxML (Stamatakis, 2014) is one of the most widely used programs in phylogenetic analysis, implementing extremely fast algorithms to analyze large datasets using maximum likelihood. Despite the undisputed efficiency of RAxML, the program is only available through a command-line interface. This requires users to be familiar with the shell environment and to navigate through the ever-growing number of commands implemented in the program, which may exclude many potential users without such experience. RaxmlGUI (Silvestro and Michalak, 2012) is a graphical interface intended to facilitate phylogenetic analyses using RAxML by providing a graphical front-end to help users set up their analysis. Although this interface has been widely used, there are many areas of improvement in terms of accessibility, usage and performance.

Here, we present the first public beta version of raxmlGUI 2.0, a complete rewrite of the raxmlGUI program. This version brings a new cross-platform design, novel functionalities and a seamless integration with RAxML 8.2. Similarly to its predecessor, raxmlGUI 2.0 is designed to be easy to use, providing the user with an intuitive interface with access to the model settings required to setup and run a phylogenetic analysis. Here we describe the available options of raxmlGUI 2.0-beta and outline the upcoming features that will be supported in the official release of raxmlGUI 2.0.

Main features

The program comes with pre-compiled integrated versions of RAxML for the major operating systems (MacOS, Windows, Linux), including the PTHREADS and SSE3 versions (Stamatakis, 2014) allowing the user to run faster analyses using parallel computing, when multiple CPUs are available. RaxmlGUI 2.0 is structured in two parts

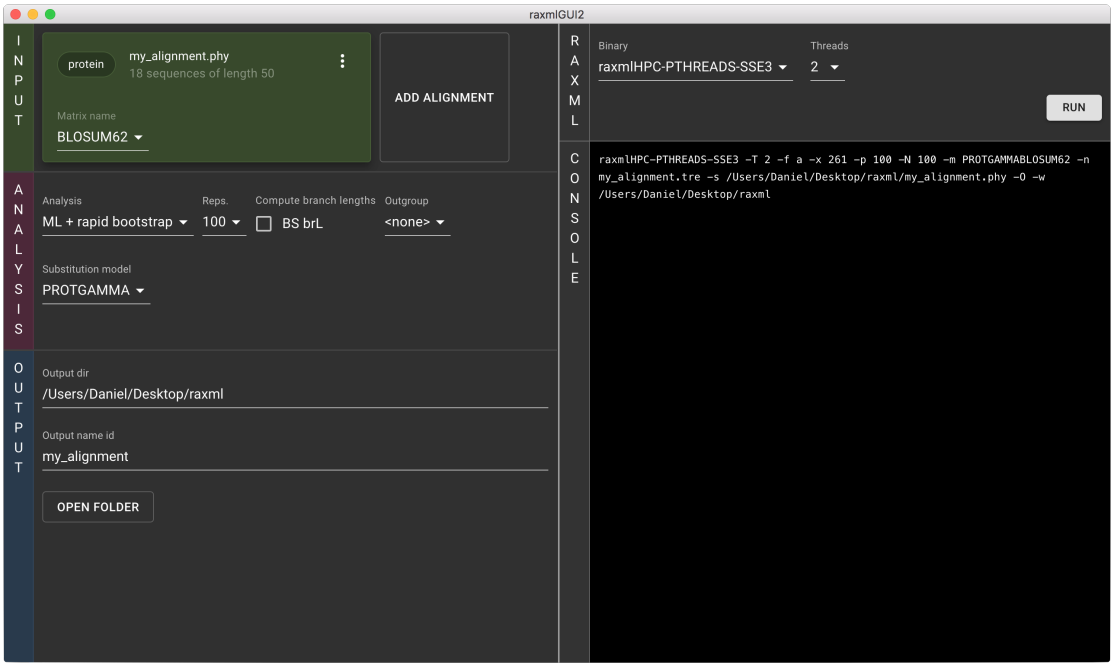

Figure 1: The raxmlGUI 2.0 interface, upon loading an alignment file with amino acid sequences. The INPUT panel provides options to load new alignments and create a concatenated file and to specify partition-specific substitution matrix. The ANALYSIS panel provides options to specify the type of analysis, evolutionary models and outgroup selection. The OUTPUT panel gives easy access to the folder with the input files and a list of output files that appears upon completing the analysis. On the right side of the window the user can select the version of RAxML, start the analysis, and visualize the RAxML output.

(Fig. 1), providing on the left all the commands and options to load input files, set up the analysis, define substitution models and partitions, among other features. On the right panel, it provides options to choose the RAxML version and start the analysis. A RAxML console is integrated in the GUI showing the progress of the analysis, the commands used to launch the analysis and all the screen output produced by RAxML.

Basic setup

RaxmlGUI 2.0 supports alignment files in two formats: extended PHYLIP and FASTA (example files are available in the program's repository). Upon loading an alignment, the program parses the names attributed to each sequence (e.g. the species name) and creates a list of taxa in the *Outgroup* menu button, which can be used to

root the tree based on a user-defined outgroup (note that maximum likelihood trees can always be re-rooted after the analysis using tree-viewing software such as FigTree (Rambaut, 2012)).

Phylogenetic analyses can be run based on different types of data: nucleotide sequences (DNA, RNA), amino acid sequences, discrete binary and multi-state characters (e.g. used for descriptions of morphological data). Since each data type requires a specific class of substitution models, raxmlGUI automatically recognizes the data type from the loaded input file and provides the user with a drop-down menu showing all the substitution models compatible with the alignment.

Analytical pipelines

The default analysis includes a maximum likelihood search of the best tree, followed by a bootstrap analysis (option *ML + rapid bootstrap*). Bootstrap support values are then drawn onto the maximum likelihood tree. After loading the alignment file and setting up the preferred substitution model, launching the default analysis only requires hitting the *Run* button on the right panel. Other options are available on the ANALYSIS panel to set the number of bootstrap pseudo-replicates. The analysis progress can be monitored on the CONSOLE panel of raxmlGUI 2.0. When the analysis is completed a list of output files will be available on the OUTPUT panel. Clicking on the file names will open the files in the user's default program (e.g. FigTree for tree files). The most important output of this analysis is named "RAxML_bipartitions.*input*.tre" (where *input* is by default the file name of the alignment) and includes the maximum likelihood tree topology and branch lengths with labels reporting the bootstrap scores for each node (bipartition) in the tree. All output files are by default saved in the same directory of the input file.

Other types of analysis are available in raxmlGUI 2.0. Some analyses integrate multiple calls to RAxML to simplify the user experience in a single pipeline. For instance, the *ML + thorough bootstrap* option launches a sequence of three RAxML calls

to 1) infer the maximum likelihood tree through a user-defined number of independent searches; 2) run a user-defined number of thorough non-parametric bootstrap replicates; and 3) draw the bootstrap support values onto the maximum likelihood tree.

Automatic concatenation of alignments

An important feature of raxmlGUI 2.0 is the automated concatenation and partitioning of alignments, which simplifies the analysis of multiple genes or combination of different data types, e.g. amino acids sequences and morphological data. After loading the first alignment, the user can add new ones to concatenate them into a single analysis. Upon loading additional alignments, raxmlGUI 2.0 performs the following tasks:

- Parse the data to determine the data type (nucleotides, amino acids, multistate)
- Parse the taxa names to make sure the concatenation of sequences occurs across matching taxa even if they are listed in different order among input files
- Create a combined dataset file in the same directory as the input files
- Create a file defining the boundaries of each partition in the concatenated alignment (each alignment file is assigned a new partition) and the respective substitution matrix (for amino acid data only)
- For any mismatch between taxa of different partitions, give option to automatically create sequences of missing data in the concatenated alignment or drop taxa with missing sequences in any partition.

These features facilitate the concatenation of different alignment files and the creation of the partition files. They also reduce the probability of errors stemming from manually merging sequences by matching taxa names. Finally, raxmlGUI 2.0 also facilitate the generation of sparse matrices resulting from the combination of alignments with different and only partly overlapping taxonomic coverage.

Upcoming features

Upcoming updates of raxmlGUI will include full support of the latest version of RAxML-ng (Kozlov et al., 2019), which provides improved performance for very large datasets.

Additional features will allow users to enforce topological tree constraints, to compute Robinson-Foulds distances (Pattengale et al., 2007), and to enable more flexibility setting substitution models and dataset partitions.

Implementation

RaxmlGUI 2.0 is built with Electron, a framework for creating cross-platform desktop applications using web technologies like JavaScript, HTML, and CSS. The user interface is built with Material-UI, a React UI framework with components that implement Google's Material Design.

The Electron base improves the portability and compatibility across platforms and operating systems compared to the previous version of raxmlGUI that uses an obsolete Python 2.x codebase. The installation is extremely simple and does not require any additional external libraries or dependencies, nor does it require admin rights on the machine.

On machines featuring multiple CPUs (i.e. most desktop and laptop computers) the GUI allows users to easily use RAxML's powerful parallel computing, which can drastically speed up the analyses. RaxmlGUI 2.0 includes pre-compiled versions of the PTHREAD version of RAxML and a dropdown menu button to specify the desired number of CPUs allocated for the analysis.

Availability and users' feedback

A public beta version of raxmlGUI 2.0 is available at antonellilab.github.io/raxmlGUI/ The program is open source and licensed under a

GNU Affero General Public License v3 (AGPL-3.0). We encourage users to report any issues, feature requests, and general feedback either as issues on GitHub github.com/AntonelliLab/raxmlGUI/issues (this requires a GitHub account) or by email at raxmlgui.help[at]gmail.com.

References

Kozlov, A. M., D. Darriba, T. Flouri, B. Morel, and A. Stamatakis. 2019. Raxml-ng: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics Page doi:10.1093/bioinformatics/btz305.

Pattengale, N. D., E. J. Gottlieb, and B. M. Moret. 2007. Efficiently computing the robinson-foulds metric. Journal of Computational Biology 14:724–735.

Rambaut, A. 2012. Figtree v. 1.4: available from https://github.com/rambaut/figtree.

Silvestro, D. and I. Michalak. 2012. raxmlGUI: a graphical front-end for RAxML. Organisms Diversity and Evolution 12:335–337.

Stamatakis, A. 2014. Raxml version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.