APPLICATION

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

Daniel Edler[1,2,3] (ID) | Johannes Klein[2] (ID) | Alexandre Antonelli[2,3,4,5] (ID) |
Daniele Silvestro[2,3,6,7] (ID)

[1]Integrated Science Lab, Department of Physics, Umeå University, Umeå, Sweden; [2]Gothenburg Global Biodiversity Centre, Gothenburg, Sweden; [3]Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden; [4]Royal Botanic Gardens, Kew, Richmond, UK; [5]Department of Plant Sciences, University of Oxford, Oxford, UK; [6]Department of Biology, Swiss Institute of Bioinformatics, University of Fribourg, Fribourg, Switzerland and [7]Swiss Institute of Bioinformatics, Fribourg, Switzerland

Correspondence
Daniel Edler
Email: daniel.edler@umu.se

Daniele Silvestro
Email: daniele.silvestro@unifr.ch

Abstract

1. raxmlGUI is a graphical user interface to RAxML, one of the most popular and widely used softwares for phylogenetic inference using maximum likelihood.

2. Here we present raxmlGUI 2.0, a complete rewrite of the GUI which seamlessly integrates RAxML binaries for all major operating systems with an intuitive graphical front-end to setup and run phylogenetic analyses.

3. Our program offers automated pipelines for analyses that require multiple successive calls of RAxML, built-in functions to concatenate alignment files while automatically specifying the appropriate partition settings, and one-click model testing to select the best substitution models using ModelTest-NG. In addition to RAxML 8.x, raxmlGUI 2.0 also supports the new RAxML-NG, which provides new functionality and higher performance on large datasets.

4. raxmlGUI 2.0 facilitates phylogenetic analyses by coupling an intuitive interface with the unmatched performance of RAxML.

KEYWORDS

bioinformatics, evolutionary biology, molecular biology, phylogenetics, software

1 | INTRODUCTION

Phylogenetic inference is a keystone in evolutionary biology research, providing the foundations for tackling a wide range of questions, from population dynamics to taxonomy of higher taxa (Felsenstein, 2003). RAxML is one of the most widely used programs in phylogenetic analysis, implementing extremely fast algorithms to analyse large datasets using maximum likelihood (Stamatakis, 2014). 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 & Michalak, 2012) is a graphical interface intended to facilitate phylogenetic analyses using RAxML by providing a graphical front-end to help users setup 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 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 both RAxML 8.2 and RAxML-NG, the new RAxML Next Generation (Kozlov et al., 2019). Similar to its predecessor, raxmlGUI 2.0 is designed to be easy to use as a cross-platform stand-alone program that does not require the installation of additional software or an internet connection. It provides the user with an intuitive interface with access to the model settings required to setup and run a phylogenetic analysis, and can handle datasets comprising up to many thousand terminals (Lemoine et al., 2018). The GUI additionally includes an option to choose the best fitting substitution model and provides a number of automated options to parse, concatenate and partition alignments and to run analytical pipelines combining multiple RAxML calls. raxmlGUI 2.0 targets a wide user base ranging from beginners to advanced phylogeneticists seeking an easy access to the state-of-the-art analytical tools implemented in RAxML and RAxML-NG.

2 | METHODS

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. Pre-compiled versions of RAxML-NG are provided for MacOS and Linux. A Windows version will be added when available from the RAxML-NG development team.

raxmlGUI 2.0 is structured into five different sections, INPUT, ANALYSIS, OUTPUT, RAXML and CONSOLE (Figure 1). The left panel with the three first sections lets the user load input files, setup the analysis, define substitution models and partitions, choose output path, among other features. The right panel lets the user select RAxML version, see and run the command resulting from the input on the left panel, and see the progress and output from RAxML in the integrated console.

2.1 | Basic setup

raxmlGUI 2.0 supports alignment files in different formats commonly used in phylogenetic analyses: extended PHYLIP, FASTA, NEXUS and MEGA (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
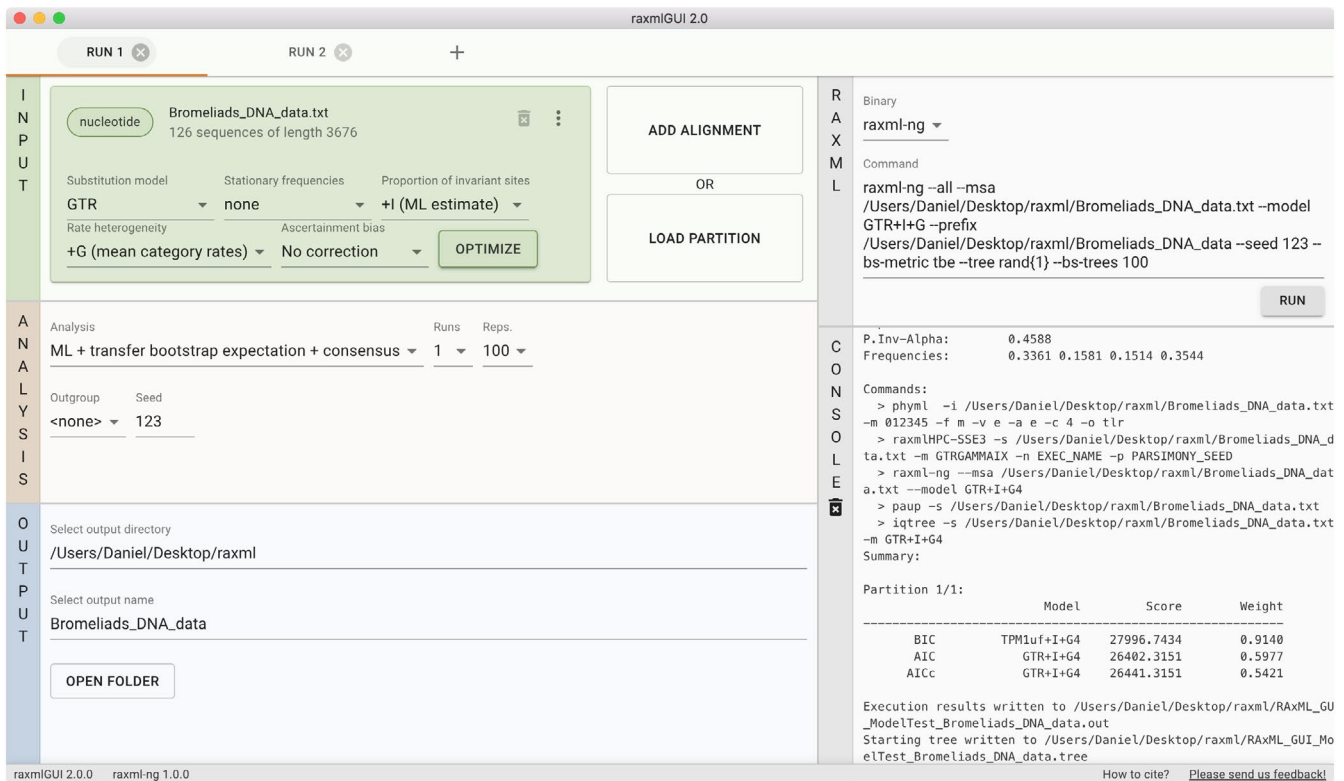

FIGURE 1 The raxmlGUI 2.0 interface, upon loading an alignment file with nucleotide sequences. In the left panel, the input section provides options to load new alignments, create a concatenated file for partitioned datasets, and to specify partition-specific substitution matrices. The optimize button allows the user to perform model testing and automatically specify the best-fitting substitution model. The analysis section provides options to specify the type of analysis and outgroup selection. The output section gives easy access to the folder with the input files and a list of output files that appears upon completing the analysis. In the right panel, the user can select the version of RAxML (RAxML-ng shown in the figure), start the analysis and visualize the output. A tab bar on the top of the window allows users to easily setup multiple runs and switch between them

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 2.0 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.

2.2 | Analytical pipelines

Analytical pipelines readily implemented in raxmlGUI 2.0 include a maximum likelihood search of the best tree, followed by a bootstrap analysis. Bootstrap support values are then drawn onto the maximum likelihood tree. After loading the alignment file and setting up the preferred substitution model (options for model testing directly from raxmlGUI 2.0 are described below), 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 in the console section of raxmlGUI 2.0. When the analysis is completed, a list of output files will be available in the output section. 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.

Several 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, in a simple click, a sequence of three RAxML calls to (a) infer the maximum likelihood tree through a user-defined number of independent searches (b) run a user-defined number of thorough non-parametric bootstrap replicates and (c) draw the bootstrap support values onto the maximum likelihood tree.

2.3 | Automatic concatenation of alignments and partitions

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, for example, amino acid 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.
- 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.
- Set default partitions for the new alignments and re-compute the concatenated partition.

These features facilitate the concatenation of different alignment files, the creation of the partition files and the generation of sparse matrices resulting from the combination of datasets with different and only partly overlapping taxonomic coverage. These tools also reduce the probability of errors stemming from manually merging sequences by matching taxa names. Additionally, raxmlGUI 2.0 provides an intuitive interface to create partitions within a single alignment file, including the possibility to specify codon based evolutionary models for coding nucleotide sequences (Figure 2). Finally, the user can load their own partition files, which must be provided in a RAxML compatible format (Figure 1).

2.4 | Support for both RAxML 8.x and RAxML-NG

In addition to RAxML 8.x, raxmlGUI 2.0 adds support for RAxML Next Generation (Kozlov et al., 2019), which provides new options and improved performance for very large datasets, which are typical for the analyses of genomic data. Among the novel methods implemented in RAxML-NG, and available through raxmlGUI 2.0, is the Transfer Bootstrap Expectation algorithm to quantify topological support for a tree (Lemoine et al., 2018). This algorithm has been shown to outperform the traditional bootstrap analysis (Felsenstein, 1985) when applied to large phylogenetic trees (thousands of tips). The user can select which version of RAxML they want to run from the GUI, and the available settings are automatically updated for the specific version. For guidelines of which RAxML version to use for particular objectives and datasets, please refer to Kozlov et al. (2019).

2.5 | Model testing

One of the advantages of RAxML-NG over RAxML is its increased range of available substitution models for nucleotide and amino acid data. This feature also allows users to define different substitution models for each partition, for example, when analysing concatenated genes. To facilitate the use of these features, we implemented a model testing feature in raxmlGUI 2.0 that allows the user to select the best substitution model based on the corrected
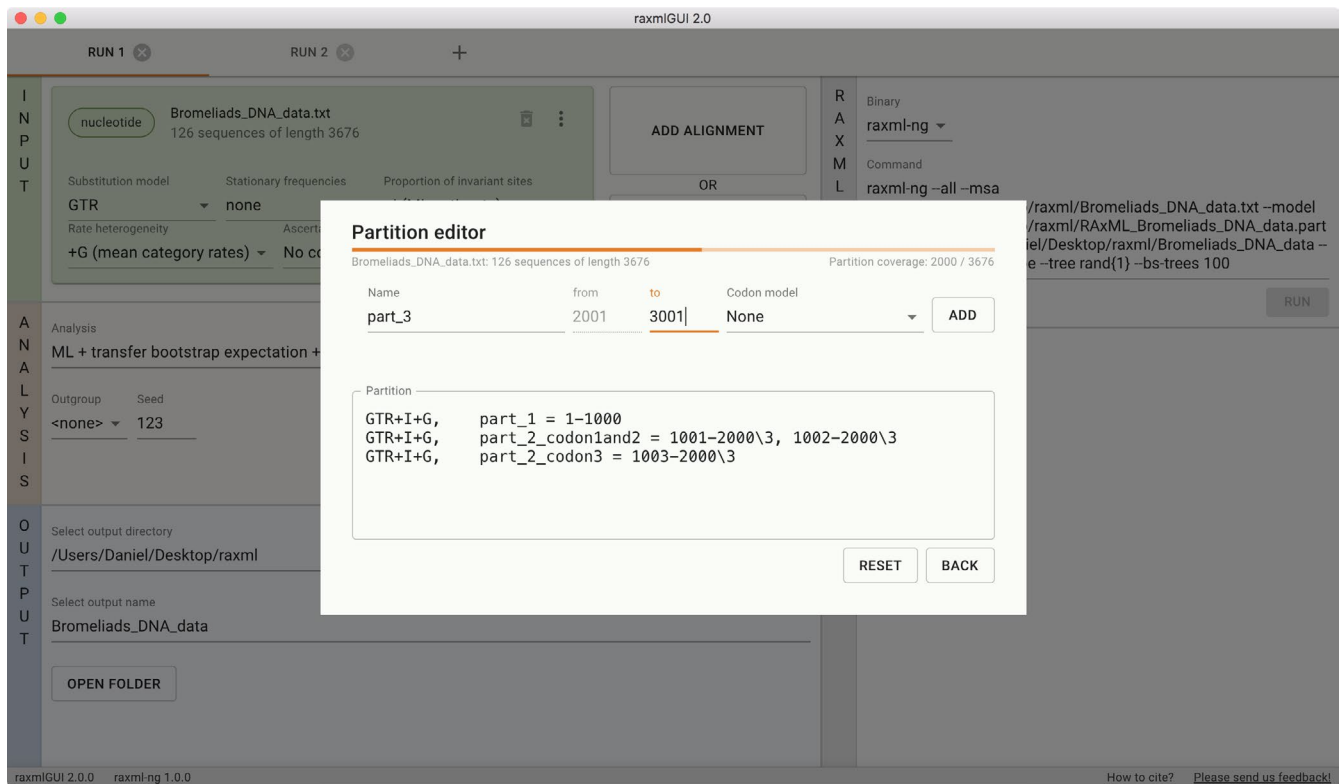
FIGURE 2 The raxmlGUI 2.0 partition editor. A graphical user interface makes it easy to define a partition for individual alignments

Akaike Information Criterion (AICc; Burnham & Anderson, 2002). Model testing is carried out using the program ModelTest-NG (Darriba et al., 2019), and is seamlessly integrated within raxml-GUI 2.0 through the OPTIMIZE button (Figure 1). The test can be run separately for each partition and the best model will be specified automatically for the following analysis. As for RAxML-NG, ModelTest-NG is currently provided for MacOS and Linux, whereas Windows support will be added as soon as a compatible version is made available by the ModelTest-NG development team.

2.6 | Performance and implementation

There is no performance difference between running RAxML on the command line and running it from the GUI as raxmlGUI 2.0 just forwards all settings as parameters to the command line version of RAxML and runs that as a separate process. raxmlGUI 2.0 also supports a tabbed interface for running multiple analyses in parallel (Figure 1).

raxmlGUI 2.0 is built with Electron (Github Inc., 2020), a framework for creating cross-platform desktop applications using web technologies such as JavaScript, HTML and CSS. The user interface is built with Material-UI (Material-UI, 2020), a React (Facebook Inc., 2020) user interface framework with components that implement Google's Material Design (Google, 2020). 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.

3 | CONCLUSION

We presented a graphical interface providing an intuitive and user-friendly access to the high-performance phylogenetic software RAxML and RAxML-NG, without compromising performance. Our implementation allows students, professionals and researchers to use the latest, state-of-the-art methods to build robust phylogenetic hypotheses, irrespective of their computing skills. We hope research and teaching in different fields involving phylogenetic inference—from evolutionary biology to taxonomy, from drug discovery to epidemiology—can benefit from using our program.

AUTHORS' CONTRIBUTIONS

D.S. conceived the idea; D.E, J.K. and D.S. designed the interface; D.E. and J.K. implemented the ideas and design; D.E., J.K. and D.S. led the writing of the manuscript; D.E., J.K., D.S. and A.A. contributed critically to the drafts and gave final approval for publication.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13512.

DATA AVAILABILITY STATEMENT

raxmlGUI 2.0 is available for download for Windows, macOS and Linux at https://antonellilab.github.io/raxmlGUI/, with links to example files for testing. The program is open source and licensed under the GNU Affero General Public License v3.0 or later (AGPL-3.0-or-later). The source code is available on GitHub (https://github.com/AntonelliLab/raxmlGUI). The RAxML binaries included in raxmlGUI 2.0 are also available in a separate GitHub repository (https://github.com/AntonelliLab/raxmlGUI-binaries). All source code, binaries and data used in this paper are archived on Zenodo with https://doi.org/10.5281/zenodo.4065595 (Edler et al., 2020); Zenodo. We encourage users to report any issues, feature requests and general feedback either as GitHub issues at https://github.com/AntonelliLab/raxmlGUI/issues (this requires a GitHub account) or by email to raxmlgui.help@gmail.com.

ORCID

Daniel Edler (iD) https://orcid.org/0000-0001-5420-0591
Johannes Klein (iD) https://orcid.org/0000-0003-2736-8094
Alexandre Antonelli (iD) https://orcid.org/0000-0003-1842-9297
Daniele Silvestro (iD) https://orcid.org/0000-0003-0100-0961

REFERENCES

Burnham, K. P., & Anderson, D. A. (2002). *Model selection and multi-model inference: A practical information-theoretic approach* (2nd ed.). Springer.

Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., & Flouri, T. (2019). ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Molecular Biology and Evolution*, *37*, 291–294.

Edler, D., Klein, J., Antonelli, A., & Silvestro, D. (2020). raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML. v2.0.0, *Zenodo*, https://doi.org/10.5281/zenodo.4065595

Facebook, Inc. (2020). *React – A JavaScript library for building user interfaces*. Copyright (c) Facebook, Inc. and its affiliates. Retrieved from https://reactjs.org

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, *39*, 783–791.

Felsenstein, J. (2003). *Inferring phylogenies*. Sinauer Associates.

Github Inc. (2020). Electron – Build cross-platform desktop apps with JavaScript, HTML, and CSS. Copyright (c) GitHub Inc. Retrieved from https://www.electronjs.org

Google. (2020). *Material design – Material is an adaptable system of guidelines, components, and tools that support the best practices of user interface design*. Retrieved from https://material.io

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, *35*, 4453–4455.

Lemoine, F., Entfellner, J.-B.-D., Wilkinson, E., Correia, D., Felipe, M. D., De Oliveira, T., & Gascuel, O. (2018). Renewing felsenstein's phylogenetic bootstrap in the era of big data. *Nature*, *556*, 452–456.

Material-UI. (2020). *Material-UI – React components for faster and easier web development*. Copyright (c) Material-UI. Retrieved from https://material-ui.com

Rambaut, A. (2012). *Figtree v. 1.4*. Retrieved from https://github.com/rambaut/figtree

Silvestro, D., & Michalak, I. (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.

How to cite this article: Edler D, Klein J, Antonelli A, Silvestro D. raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML. *Methods Ecol Evol.* 2021;12:373–377. https://doi.org/10.1111/2041-210X.13512