Org Divers Evol (2012) 12:335–337
DOI 10.1007/s13127-011-0056-0

METHODS AND APPLICATIONS

ORGANISMS DIVERSITY & EVOLUTION

# raxmlGUI: a graphical front-end for RAxML

**Daniele Silvestro · Ingo Michalak**

Received: 29 March 2011 / Accepted: 22 August 2011 / Published online: 1 September 2011


**Abstract** With the increasing availability of molecular data, maximum likelihood approaches have gained a new central role in phylogenetic reconstructions. Extremely fast tree-search algorithms have been developed to handle data sets of ample size in reasonable time. In the past few years, RAxML has achieved great relevance in this field and obtained wide distribution among evolutionary biologists and taxonomists because of its high computational performance and accuracy. However, there are certain drawbacks with regard to its usability, since the program is exclusively command-line based. To overcome this problem, we developed raxmlGUI, a graphical user interface that makes the use of RAxML easier and highly intuitive, enabling the user to perform phylogenetic analyses of varying complexity. The GUI includes all main options of RAxML, and a number of functions are automated or simplified. In addition, some features extend the standard use of RAxML, like assembling concatenated alignments with automatic partitioning. RaxmlGUI is an open source Python program, available in a cross-platform package that incorporates RAxML executables for the main operating systems. It can be downloaded from http://sourceforge.net/projects/raxmlgui/.

**Keywords** Rapid bootstrap · Graphical user interface · Maximum likelihood · Phylogenetic analyses · Python · RAxML

## Introduction

With the advent of new techniques to obtain DNA sequences and the increasing availability of molecular data in online databases, phylogenetic analysis often becomes the bottleneck in evolutionary biology and taxonomic research. For many biologists it is a major challenge to cope with the available cutting edge programs, since most of them are command-line based and lack a graphical user interface (GUI). One of these programs, RAxML (Stamatakis 2006a), has become a standard application for maximum likelihood (ML) phylogenetic analyses in recent years (e.g., Guindon et al. 2010; Blair and Murphy 2011) and has been developed continuously (e.g., Ott and Stamatakis 2010; Stamatakis and Alachiotis 2010; Pattengale et al. 2011). Apart from its exceptional computational speed and accurate algorithms (Stamatakis et al. 2008) RAxML's most important features include multi-thread computation, options for data partitioning, substitution models for morphological and amino acid data, and for nucleotide secondary structures. The analyses are normally launched from shell through a single string of commands, defining all the settings and input files. Web interfaces are available, e.g., the "Blackbox"

Daniele Silvestro and Ingo Michalak contributed equally

D. Silvestro · I. Michalak
Department of Botany and Molecular Evolution,
Research Institute Senckenberg,
Senckenberganlage 25,
60325, Frankfurt am Main, Germany

D. Silvestro · I. Michalak
Biodiversity and Climate Research Centre (BiK-F),
Senckenberganlage 25,
60325, Frankfurt am Main, Germany

D. Silvestro (✉) · I. Michalak (✉)
Department of Diversity and Evolution of Higher Plants, Institute of Ecology, Evolution and Diversity, Goethe University Frankfurt,
Senckenberganlage 25,
60325, Frankfurt am Main, Germany
e-mail: daniele.silvestro@senckenberg.de
e-mail: ingo.michalak@senckenberg.de

(Stamatakis et al. 2008), which, however, lacks several options in the analysis settings.

To facilitate the use of RAxML, we have developed raxmlGUI – a program that allows the user to set up and run ML analyses through a graphical interface. In raxmlGUI, a number of options and functions are automated (e.g., checking for identical sequences, or gap-only characters) and simplified (e.g., model and outgroup selection, excluding sites, and partitioning a matrix). Some features, like assembling concatenated data sets with automatic partitioning, extend the standard use of RAxML.

## Usage and implemented features

### System requirements

RaxmlGUI is a Python program based on the module Tk, which is a standard library included in the main Python package (2.5 or higher). The program is cross-platform and incorporates by default RAxML executables compiled for Windows, Mac OSX (64bit), and Linux in single-core and multithread versions. Additional versions can be integrated easily using the Preferences panel.

### Loading a file

Alignments in PHYLIP format can be loaded in raxmlGUI and are automatically checked for their compatibility with RAxML. The alignment can be sequential or interleaved, details on the format can be found in Stamatakis (2008). If identical sequences or gap-only characters are detected, the user can exclude them, generating a reduced data set that is loaded automatically. A combined analysis of separate alignment files can be performed through the interface by choosing "Add alignment", which concatenates the data sets into a single matrix and defines the appropriate partitioning settings (see below).

### Setting up an analysis

When an alignment is loaded, raxmlGUI recognizes the data type and provides the possibility of choosing from among compatible model settings. The available substitution models include generalized time reversible (GTR) for nucleotides, symmetric, Markovian, and ordered models for morphological characters; and several amino-acid models with the option to use empirical base frequencies. Each substitution model is complemented with GAMMA (Yang 1994) or CAT (Stamatakis 2006b) models of rate heterogeneity among sites, with or without a proportion of invariable sites. For nucleotide alignments, a secondary structure file can be passed to define stem and pseudo-knot regions.

By default, the ML analyses generate unrooted phylogenies. The trees can be rooted with a single-taxon outgroup selected from a list of taxa available in the program's toolbar, whereas a multi-taxon outgroup can be defined through the menu. A topological constraint can be enforced by defining monophyletic groups from within raxmlGUI or by loading a Newick-formatted binary or multifurcating tree-file.

Portions of the alignment can be excluded from the analysis using the "Exclude sites" option. This will produce a new alignment file that is automatically set as the input file and checked for identical sequences.

### Analysis options

Three different ML analyses can be set up through raxmlGUI:

(1) Maximum likelihood reconstruction using the rapid hill-climbing algorithm (Stamatakis et al. 2007) (RAxML option "-f d"; "ML search" option in the GUI)
(2) Rapid bootstrap analysis (Stamatakis et al. 2008) and search for a best-scoring ML tree (RAxML option "-f a").
(3) Thorough bootstrap analysis (RAxML option "-b"), followed by a ML search. Subsequently, BS support values are drawn on the best-scoring ML tree (RAxML option "-f b").

The number of independent ML searches and BS replicates can be set using the program's toolbar. As an alternative to the predefined number of BS replicates, different "bootstopping" options are available (Pattengale et al. 2010), which automatically stop the bootstrap run after the necessary number of replicates.

RaxmlGUI further incorporates RAxML's options to generate consensus trees, compute per site log-likelihoods with output compatible with the software CONSEL (Shimodaira 2001), and to compute Robinson-Foulds pairwise distances (Robinson and Foulds 1981) between trees.

### Partitioned models

An important feature of RAxML is that it supports concatenated alignments in which different data types, e.g., morphological and molecular data, are combined. RaxmlGUI implements a tool to facilitate the set-up of the partitions and their respective substitution models. In addition, an option to assemble a partitioned data set is implemented. Individual alignment files can be loaded subsequently to produce a concatenated matrix while setting the appropriate model options. If the "per-partition brL"-box is checked, the

branch lengths will be calculated independently for each partition.

## Conclusion and scope

RaxmlGUI is intended to accelerate and simplify the usage of RAxML, enabling an interactive control of all its major features (as of RAxML version 7.2.8). The graphical interface is designed to be self-explanatory and to make its use very intuitive. In addition, a detailed built-in help file is available. Through the implementation of multi-thread versions of RAxML, the GUI enables the optimal utilization of the available computational resources, which is particularly important given the continuing increase in the size and complexity of molecular data sets. The scope of raxmlGUI is to extend RAxML's target group beyond scientists with high computer expertise, and to enhance its usability in teaching.

**Acknowledgments** We are grateful to Alexandros Stamatakis, for support and suggestions. Furthermore, we thank Gaëlle Bocksberger, Fernando Fernandez, Rafael Louzada, and Jan Schnitzler for beta-testing and aid. This study was supported financially by the German Research Foundation (DFG, projects ZI 557/6-2, ZI 557/7-1, SCHU 2426/1-1, SCHU 2426/2-2), and the research funding programme "LOEWE Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz" of Hesse's Ministry of Higher Education, Research, and the Arts.

The authors declare that they have no conflict of interest.

## References

Blair, C., & Murphy, R. W. (2011). Recent trends in molecular phylogenetic analysis: where to next? *Journal of Heredity, 102*, 130–138.

Guindon, S., Duyfayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology, 59*, 307–321.

Ott, M., & Stamatakis, A. (2010). Preparing RAxML for the SPEC MPI Benchmark Suite. In S. Wagner, M. Steinmetz, A. Bode, & M. M. Müller (Eds.), *High Performance Computing in Science and Engineering, Garching/Munich 2009, Part 8* (pp. 757–768). Berlin: Springer.

Pattengale, N. D., Aberer, A. J., Swenson, K. M., Stamatakis, A., & Moret, B. M. E. (2011). Uncovering hidden phylogenetic consensus in large datasets. *IEEE/ACM transactions on computational biology and bioinformatics*, 03 Feb. 2011. IEEE computer Society Digital Library. IEEE Computer Society, http://doi.ieeecomputersociety.org/10.1109/TCBB.2011.28

Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R. P., Moret, B. M. E., & Stamatakis, A. (2010). How many bootstrap replicates are necessary? *Journal of Computational Biology, 17*, 337–354.

Robinson, D. R., & Foulds, L. R. (1981). Comparison of phylogenetic trees. *Mathematical Biosciences, 53*, 131–147.

Shimodaira, H. (2001). Multiple comparisons of log-likelihoods and combining nonnested models with applications to phylogenetic tree selection. *Communications in Statistics - Theory and Methods, 30*, 1751–1772.

Stamatakis, A. (2006a). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics, 22*, 2688–2690.

Stamatakis, A. (2006b). Phylogenetic models of rate heterogeneity: A high performance computing perspective. In: *Proceedings of the 20th IEEE International Parallel & Distributed Processing Symposium (IPDPS2006)*, (pp. 278–286). Washington: IEEE Computer Society Press.

Stamatakis, A. (2008). *The RAxML 7.0.4 manual*. http://icwww.epfl.ch/~stamatak/index-Dateien/countManual7.0.4.php. Accessed 8 December 2010.

Stamatakis, A., & Alachiotis, N. (2010). Time and memory efficient likelihood-based tree searches on phylogenomic alignments with missing data. *Bioinformatics, 26, ISMB 2010*, i132–i139.

Stamatakis, A., Blagojevic, F., Nikolopoulos, D. S., & Antonopoulos, C. D. (2007). Exploring new search algorithms and hardware for phylogenetics: RAxML meets the IBM cell. *Journal of VLSI Signal Processing, 48*, 271–286. doi:10.1007/s11265-007-0067-4

Stamatakis, A., Hoover, P., & Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology, 57*, 758–771.

Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution, 39*, 306–314.