

siMBa—a simple graphical user interface for the Bayesian phylogenetic inference program MrBayes

Bagdevi Mishra · Marco Thines

Received: 18 August 2014 / Revised: 20 August 2014 / Accepted: 21 August 2014 / Published online: 13 September 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract MrBayes is a program that uses a Bayesian framework for inferring phylogenetic relationships. As MrBayes is a command-line-driven program, users acquainted to programs with graphical user interfaces will not find it easy to operate, especially as it requires a complex input format for the data to be analysed. We thus developed siMBa (simple MrBayes), a simple graphical user interface for MrBayes. This tool gives the user interactive control over most of the parameters and also facilitates the input of a multiple sequence alignment, as any widely used format can be used. siMBa is coded in Perl using the Tk module. Executables are provided for Windows, Linux, and Macintosh. The Perl codes, along with executables for different operating system, are freely available to download from [<http://www.thines-lab.senckenberg.de/simba>].

Keywords Bayesian inference · Graphical user interface · MrBayes · Perl · Phylogenetic analysis software · Tk

Introduction

Phylogenetic studies are needed to analyse evolutionary relationships either between individuals within a species or between individuals across different species. A phylogeny is inferred from events that happened in the past and cannot be concluded, but rather must be estimated (Brinkman and Leipe 2001), applying different statistical approaches to morphological and molecular data. At present, among the most widely used algorithms for phylogenetic analyses are Neighbor-joining (Saitou and Nei 1987; 1986), Minimum Evolution (Rzhetsky and Nei 1992), Maximum Likelihood (Felsenstein 1981), and Bayesian inference (Mau and Newton 1997; Yang and Rannala 1997; Mau et al. 1999). Several of these algorithms have been implemented in graphical user interfaces such as Seaview (Gouy et al. 2010), MEGA (Tamura et al. 2011), raxmlGUI (Silvestro and Michalak 2012) and TOPALi (Milne et al. 2009), and there are also web servers for conducting such analyses, e.g., T-REX (Makarenkov 2001), RAxML BlackBox (Stamatakis et al. 2008), CIPRES Science Gateway (Miller et al. 2010) and phylogeny.fr (Dereeper et al. 2008). A systematic collection of software developed for conducting phylogenetic analyses is available from (<http://evolution.gs.washington.edu/phylip/software.html>), where on the 20th of August 2014, 392 phylogenetic programs and 54 freely accessible web servers were listed, an amount that has been growing dynamically over the past years. One important software for phylogenetic analyses is MrBayes (Huelsenbeck and Ronquist 2001), a program for Bayesian inference using Markov Chain Monte Carlo (MCMC) methods (Metropolis et al. 1953; Hastings 1970; Geyer 1991). Bayesian inference is a powerful tool for disentangling the complexities of evolutionary processes (Huelsenbeck et al. 2001). MrBayes is widely used by the scientific community because of its accuracy and versatility. Even though a few analysis pipelines use MrBayes as one of their components

B. Mishra (✉) · M. Thines (✉)
Biodiversity and Climate Research Centre (BiK-F), Georg-Voigt-Str.
14-16, 60325 Frankfurt (Main), Germany
e-mail: bagdevi.mishra@senckenberg.de
e-mail: marco.thines@senckenberg.de

B. Mishra · M. Thines
Institute of Ecology, Evolution and Diversity, Goethe University,
Max-von-Laue-Str. 9, 60323 Frankfurt (Main), Germany

B. Mishra · M. Thines
Senckenberg Gesellschaft für Naturforschung, Senckenberganlage
25, 60325 Frankfurt (Main), Germany

M. Thines
Center for Integrative Fungal Research (IPF), Georg-Voigt-Str.
14-16, 60325 Frankfurt (Main), Germany

(Dereeper et al. 2008; Miller et al. 2010), there is no dedicated graphical user interface (GUI) available for MrBayes, yet. MrBayes is a command-line-driven software that needs complex input files, which creates a hurdle for its usage, as many beginning users are not used to specify parameters using a command line, and the rather complicated structure of the nexus format required for MrBayes.

In an effort to make MrBayes available to be used by a larger scientific community, we built a simple GUI for MrBayes (siMBa) for different operating systems such as Linux, Windows and Macintosh, where users can input multiple sequence alignments in several widely used formats, and, with a single click on the submit button, can start an analysis with default parameters to perform a Bayesian phylogenetic inference. If any change in the values of the parameters is needed, this can be accomplished by selecting parameters from the same screen, which makes the interface simple and easy to use. Along with the possibility to run MrBayes locally, siMBa also provides the option to generate a ready-to-use nexus file, with the sequence block and MrBayes block, from a multiple sequence alignment, in case the user wants to run the job on a personal server with higher computational power on which the installation of the graphical version is not possible.

Materials and methods

Development of the interface

The scripting language Perl was used to build siMBa. The Tk module of Perl was used to build the graphical interface. The program was divided into several modules for easier data handling and all of these modules were linked together with a main program, which uses the modules at different points of the analysis. The Bio-Perl module is used to convert the format of the input multiple sequence alignment (MSA) to nexus format, if the user supplies any other widely used MSA format. System-specific Perl codes were written for different operating systems. Self-contained Perl executables were built using the PAR-Packer package from Perl, and executables for computers with Windows, Macintosh, and different distributions of Linux as an operating system were produced.

Incorporation of MrBayes

The executable version of MrBayes (version 3.2.2) is incorporated in the package supplied for the MS Windows version of siMBa. For Macintosh and different distributions of Linux, MrBayes is compiled from the source code (version 3.2.2), and is supplied along with the respective version of siMBa.

Results

Description of the graphical user interface

The Graphical User Interface of siMBa is compact when opened, with just a few basic parameters to be set, such as selecting a multiple sequence alignment file, a substitution type, the number of generations to be run for the MCMC analysis, the percentage of trees sampled to be discarded (burnin), the name of the outgroup (if any) and a name to be prefixed to all of the result files. Help files of MrBayes are available for each parameter used and can be accessed from the “help” drop-down menu at the top right corner of the main interface window. The same information is also displayed when the mouse is rolled over any parameter name on the main window.

Parameter selection

Default parameter values are already filled in, in case a user wants to run analyses with the default values for all parameters. In cases where the output prefix is not defined by the user, the default prefix will be “siMBaoutput”. The “advanced options” button appends the window named “All other parameters” to the main window, which allows specific selection of most parameters available for phylogenetic analyses in MrBayes. After selecting parameter values, the “All other parameters” window minimises again, once the “Add parameters” button is activated. There is a “reset parameters” button available, in case the user would like to go back to the default values of the parameters. The “Build ready-to-run nexus file” option creates a nexus file with the information from the alignment file and the values selected for the program parameters of MrBayes. This file can be run from the MrBayes console environment with the simple command “execute <nexus file name>”.

Display of information

In the text box at the bottom of the GUI, the values of all parameters set will be displayed so that users can go through the values again before starting the analyses and re-adjust any value, if needed. This window also displays the name of the MSA file that is to be used in the analyses. The analysis starts after activating the “Run MrBayes” button, and the progress of MrBayes starts being displayed in a separate window. If at any point in time the analysis needs to be terminated, there is a “terminate analysis” button to terminate the run. Otherwise, after successful completion, a third window will be displayed with a list of files generated by MrBayes. A screenshot of the main windows is shown in Fig. 1. The contents of the files generated by

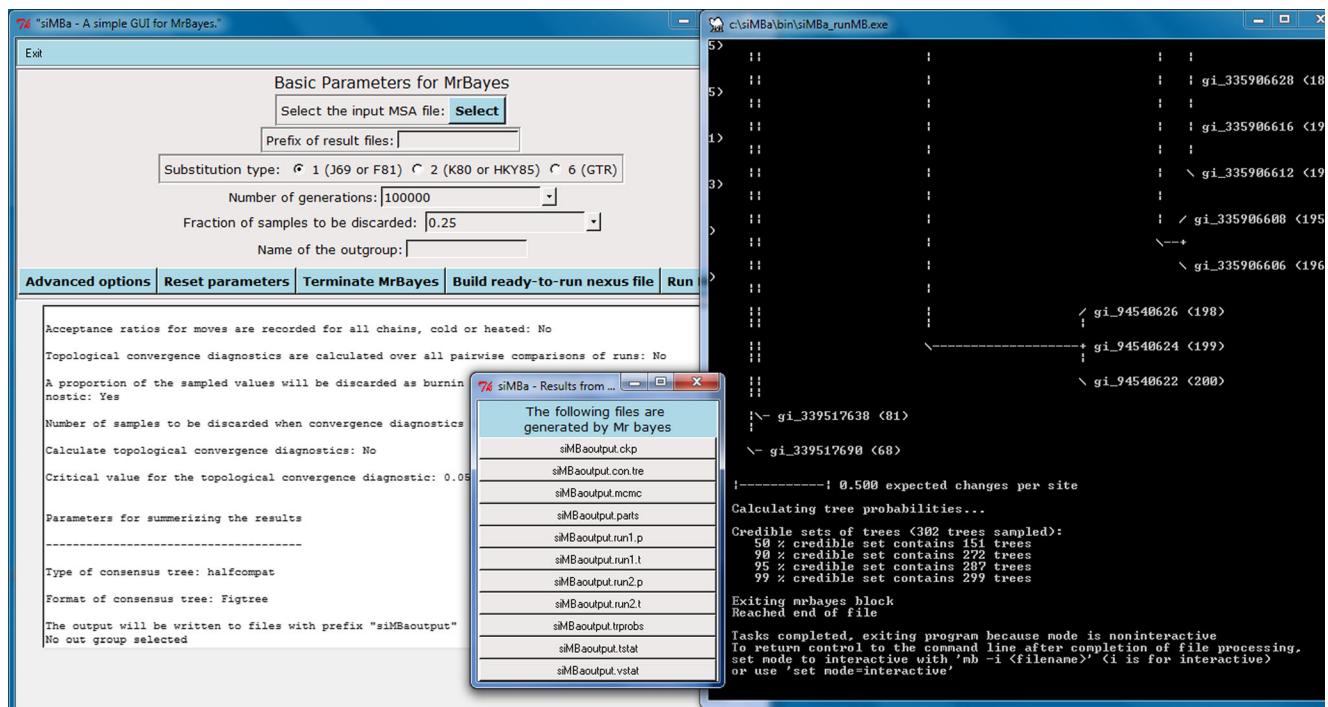


Fig. 1 Screenshot of the graphical user interface siMBa, showing basic input options (*left*) and the output windows showing the running status of MrBayes (*right*) and the list of the files generated by the analyses (*inset*)

MrBayes can be viewed by clicking on the file names. The files are stored in the same folder from which siMBa was run.

Discussion

The availability of a graphical interface (GUI) for an otherwise command-line-driven software is always advantageous, as it is easier and more intuitive to operate. Some web servers offer interfaces for MrBayes, but there are several points that pose obstacles for performing phylogenetic analyses. The MrBayes Web form from the Santos Lab (Roderic 2008) assists in creating the command block required for running MrBayes, but users cannot run MrBayes directly on the server and still need to run the MrBayes console using the block created and the nexus input format for the alignment. At the CIPRES Gateway (Miller et al. 2010), it is possible to run MrBayes on dedicated servers from CIPRES without having to use a console environment, but the procedure is multi-stepped involving multiple windows, and the users also need to prepare any input files in the nexus format used by MrBayes. On the phylogenetic.fr server (Dereeper et al. 2008), some flexibility regarding the format the alignment is available, but the number of sequences that can be input for analysis is restricted to

30, a number usually too low for detailed phylogenetic analyses. These restrictions render it difficult to carry out detailed phylogenetic analyses using MrBayes in a non-console environment.

siMBa offers an interactive control of the parameters of MrBayes that are needed to run phylogenetic analyses, rendering it simple to use. Thus, it facilitates the usage of MrBayes by removing the hurdle of a console environment and complex input files. Especially the possibility of using several widespread multiple sequence alignment formats as an input saves time and avoids file conversion to the nexus format using other software or web servers before running MrBayes. Also, the MrBayes command block, which is usually attached to the alignment file, will be created automatically after parameters have been selected in the GUI. Having no restriction for the number of input sequences, it also helps carrying out large-scale data analyses seamlessly. More features are planned to be included in future versions of siMBa, such as allowing the users to view and edit multiple sequence alignments before running the analysis allowing the usage of parameter values directly from Jmodel Test (Darriba et al. 2012), and a conversion of nexus tree files to newick format. With the simple design and easily accessible help regarding the parameters that can be used in MrBayes, we hope that siMBa will prove to be useful for application in both research and teaching.

Acknowledgments This study has been funded by the LOEWE program of the government of Hesse in the framework of the Biodiversity and Climate Research Centre (BiK-F) and Integrative Fungal Research Cluster (IPF). We are thankful to Francesco dal Grande and Deepak Kumar Gupta for input regarding the design of the work flow of MrBayes and the design of the interface, respectively. We are thankful to Claus Weiland for his help in troubleshooting of system specific issues. We are grateful to researchers at BiK-F for help with testing the software on different operating systems and to Dominik Begerow and Marc Stadler for critical assessment of the manuscript and the interface. MT conceived the study and provided input regarding the design and functionality of the interface, BM wrote the code for siMBa and organised the β-testing within the Biodiversity and Climate Research Centre, and BM and MT wrote the manuscript. The siMBa program is distributed under the GNU general public license. The program is distributed without warranty of any kind.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Brinkman FSL, Leipe DD (2001) Phylogenetic analysis. In: Baxevanis AD, Ouellette BFF (eds) Bioinformatics: a practical guide to the analysis of genes and proteins. Phylogenetic analysis. John Wiley & Sons, New York, pp 323–358
- Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772
- Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36:W465–W469
- Felsenstein J (1981) Evolutionary trees from DNA-sequences—a maximum-likelihood approach. *J Mol Evol* 17:368–376
- Geyer CJ (1991) Markov chain Monte Carlo maximum likelihood. In: Keramidas EM (ed) Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface, Fairfax Station, VA: Interface Foundation of North America, 21–24 Apr 1991, pp. 156–163
- Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 27:221–224
- Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57:97–109
- Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755
- Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Evolution—Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294:2310–2314
- Makarenkov U (2001) T-REX: reconstructing and visualizing phylogenetic trees and reticulation networks. *Bioinformatics* 17: 664–668
- Mau B, Newton MA (1997) Phylogenetic inference for binary data on dendograms using Markov chain Monte Carlo. *J Comput Graph Stat* 6:122–131
- Mau B, Newton MA, Larget B (1999) Bayesian phylogenetic inference via Markov chain Monte Carlo methods. *Biometrics* 55:1–12
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equations of state calculations by fast computing machines. *J Chem Phys* 21:1087–1091
- Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES science gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop, New Orleans, LA, 14 Nov 2010, pp 1–8
- Milne I, Lindner D, Bayer M, Husmeier D, McGuire G, Marshall DF, Wright F (2009) TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multicore desktops. *Bioinformatics* 25:126–127
- Roderic DM (2008) The Santos Lab. University of Glasgow. <http://www.auburn.edu/~santos/mrbayesform.htm>. Accessed 30 Mar 2014
- Rzhetsky A, Nei M (1992) A simple method for estimating and testing minimum-evolution trees. *Mol Biol Evol* 9:945–967
- Saitou N, Nei M (1986) The neighbor-joining method—a new method for reconstructing phylogenetic trees. *Jpn J Genet* 61:611
- Saitou N, Nei M (1987) The neighbor-joining method—a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Silvestro D, Michalak I (2012) raxmlGUI: a graphical front-end for RAxML. *Org Divers Evol* 12(4):335–337
- Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol* 57:758–771
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731–2739
- Yang ZH, Rannala B (1997) Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. *Mol Biol Evol* 14: 717–724