raxmlGUI

version 1.3

manual

Daniele Silvestro and Ingo Michalak

Senckenberg Research Institute

Biodiversity and Climate Research Centre (BiK-F)

Frankfurt am Main, Germany

raxmlgui.help@gmail.com

December 2012

Silvestro, D. and Michalak, I. 2012. raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity & Evolution* **12**, 335–337. doi: 10.1007/s13127-011-0056-0

Contents

Introduction

RaxmlGUI is a python application, which provides a user friendly front-end for RAxML (Stamatakis 2006b) for Maximum Likelihood based phylogenetic analyses. The GUI interacts with RAxML executables, which are incorporated in the package, enabling the user to select input files, set the parameters and run ML analyses, such as phylogenetic reconstructions or ancestral state reconstructions with only a few mouse clicks. 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, setting topological constraints and partitioning a matrix). Some features extend the usage of RAxML, e. g. assembling concatenated datasets with automatic partioning, and providing analyses pipelines e. g. bootstrapping followed by computing a consensus tree, or a fast tree search followed by branch lengths estimation and computing of SH-like support values.

The GUI is meant to simplify the usage of RAxML, nevertheless it is strongly recommended to get familiar with the RAxML manual (Stamatakis 2008) and the "hands-on session"[1] on the Exelixis page, to be aware of contents and intent of input and output files.

Requirements

RaxmlGUI runs under Mac, Windows and Linux operating systems. The GUI automatically determines the operating system when started for the first time, and selects the respective RAxML executable.

Python 2.5 or higher is required, but please note, that **Python 3 is not supported**. If you install the additional python library DendroPy[2] (Sukumaran & Holder 2010) it is possible to import from and export to NEXUS files (Maddison et al. 1997).

You can launch the application with a double click on the file "raxmlGUI.py" under Windows (during the installation of Python it should be set as the default application for the ".py" extension). Under Unix (Mac OS and Linux) the program is launched by browsing to its directory via shell, and typing "`./raxmlGUI.py`" or "`python raxmlGUI.py`". A Mac OSX application is also provided, that can be run with a double click.

Avoid special characters (like diacritics) and punctuation other than dots (" . ") and underscore ("_") in the (path) names of the raxmlGUI folder and all input files!

Preferences

The preferences panel provides the possibilities to change default settings and select a RAxML executable. As default you can choose among a "normal" (Windows) or a SSE3 (Mac OSX; Berger & Stamatakis 2010) and a multithread version (Ott et al. 2007). If you choose the multithread version the number of threads used by RAxML can be set. You can use the "+" button to add additional versions of RAxML (e. g. newer binaries[3] or those compiled for special needs). The chosen binary is then automatically copied to the application contents folder "/raxml". Old or unused versions can be removed with "-".

[1] http://sco.h-its.org/exelixis/hands-On.html

[2] http://packages.python.org/DendroPy/

[3] You can find the latest source code on Alexis Stamatakis' git repository or on Simon Berger's Win32 branch.

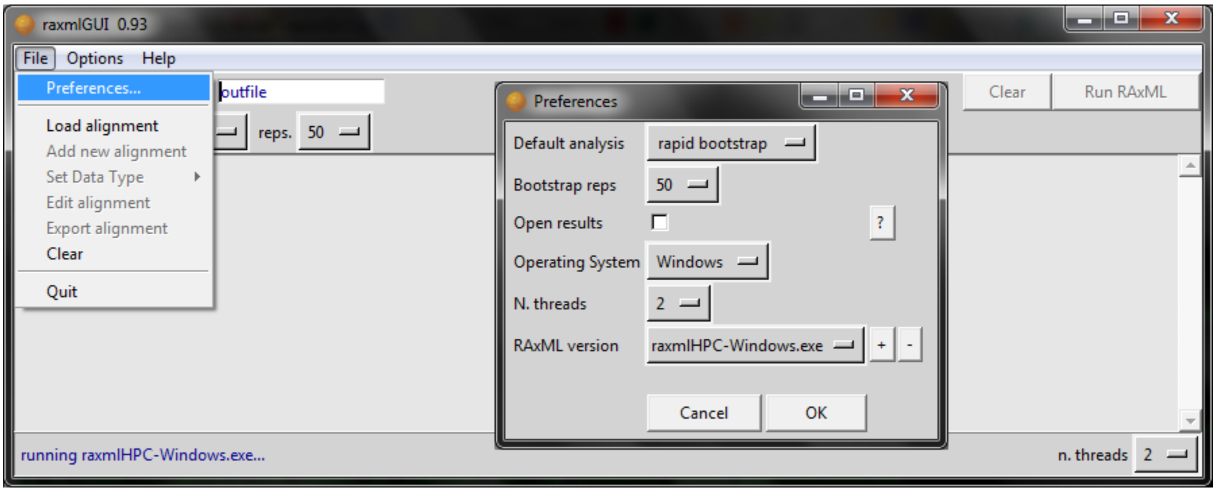

Figure 1: The preferences panel

If the box "Open results" is checked, the resulting tree (e. g. best-scoring ML tree with bootstrap values) will be opened with the default application for tree files[4] (.tre extention), if you have set one.

Please note that changes made in the preferences panel become effective only after re-starting the program.

Loading a file

With the button "Load alignment" you can load a data set as input for the RAxML analysis, which must be in PHYLIP format[5]. A FASTA or NEXUS file (Maddison et al. 1997) can be loaded with the menu option "Import FASTA/NEXUS file". The latter will be converted to a PHYLIP file for which you are asked to specify a path and file name. For converting NEXUS to PHYLIP the additional python library DendroPy (Sukumaran & Holder 2010) is needed. In case it is not yet installed, you will be guided through the installation process[6]. **Please note, that blanks and special characters in taxon names in the imported file will cause errors, as they are not allowed in PHYLIP format**, please rename your taxa according to PHYLIP limitations[7].

The data type of the loaded file is determined automatically. If the file contains only one data type (i. e. nucleotides, amino acid, binary, or multistate), the alignment will be checked for readability (through the RAxML option "`-f c`". A warning appears in case identical sequences and/or gap-only characters are detected: you can choose to run the analyses on either the original or the reduced data set. Note that if you want to exclude sites, partition the matrix and/or add additional data sets to a combined alignment, you should retain the original file, since through this option taxa might be removed and/or coloumn numbers might change.

[4] Widely used tree visualizers are Dendroscope (Huson et al. 2007), FigTree (Rambaut 2006), and Treeview (Page 1996). For a comprehensive list of tree visualizing and editing software see http://bioinfo.unice.fr/biodiv/Tree_editors.html.

[5] For an example see Appendix A

[6] If you agree to install the library, the latest source code from Jeet Sukumaran's git repository is downloaded as a zip folder. You will need to unzip this folder and pass its path to raxmlGUI, when it asks for it. You can follow the installation process in the terminal/console window. You will need to restart raxmlGUI to make use of DendroPy.

[7] Not allowed characters in PHYLIP are: , ') (: ;] [

Analysis

Analysis settings

Seven different main analyses can be carried out through raxmlGUI:

1. "**Fast tree search**" – very fast, superficial tree search (RAxML option "`-f E`") followed by optional computations of branch lengths ("`-f e`") and SH-like support values ("`-f J`"; Shimodaira & Hasegawa 1999)[8]. The analysis result is comparable to FastTree (Price et al. 2010) outputs, but is expected to yield better likelihood scores (A. Stamatakis, pers. comm.).

2. "**ML search**" – Maximum likelihood reconstruction using the rapid hill-climbing algorithm ("`-f d`"; Stamatakis et al. 2007), optionally followed by the computation of SH-like support values (see 1.), which will be plotted on the single best-scoring tree[8]. To combine the resulting trees of independent ML searches in one file, check the box "combined output".

3. "**ML + rapid bootstrap**" (default) – Rapid bootstrap analysis and search for a best-scoring Maximum Likelihood tree (equivalent to 2.) ("`-f a`"; Stamatakis et al. 2008). The bootstrap values are reported on the ML tree.

4. "**ML + thorough bootstrap**" – Thorough bootstrap analysis ("`-b`"), followed by a ML search (2.). The bootstrap support values are drawn on the most likely tree ("`-f b`").

5. "**Bootstrap + consensus**" – Rapid bootstrap analysis ("`-x`") and a subsequent majority rule consensus tree calculation from all bootstrap trees[8] ("`-J MR`").

6. "**Ancestral states**" – Compute marginal ancestral states based on a user provided rooted tree and a character matrix ("`-f A`").

7. "**Pairwise distances**" – Compute distances for all taxa pairs in the data set ("`-f x`"). As default a MP starting tree will be calculated, alternatively you can provide a user defined tree. This function is only available for GAMMA models.

Depending on the kind of alignment loaded you can choose the substitution model (GTR, BIN, MULTI, or PROT) with GAMMA[I] (Yang 1994), or CAT[I] (Stamatakis 2006a) rate heterogeneity, or without rate heterogeneity ("`-V`"). For large data sets it is possible to select the RAxML option "`-F`" from the menu to reduce the used memory (works best in combination with CAT[I]).

If the file contains **amino acid** data, you can specify the substitution model[9] you want to apply, and whether base frequencies should be determined empirically (note that the GTR substitution models always rely on empirical frequencies). If the data type in the file is **multistate** you can choose between GTR, Ordered, and MK substitution model.

You can set the **number of bootstrap replicates** with the option button "reps". You can choose a predefined number, set a user defined value, or select options of automatic "bootstopping" (Pattengale et al. 2010) according to different methods such as majority

[8]Such trees can be converted to a FigTree compatible format using the menu option "Convert to FigTree".

[9]Available substitution models for amino acids are: DAYHOFF, DCMUT, JTT, MTREV, WAG, RTREV, CPREV, VT, BLOSUM62, MTMAM, LG, MTART, MTZOA, PMB, HIVB, HIVW, JTTDCMUT, FLU, GTR and GTR_UNLINKED. For the references of these see Appendix D.

rule tree based criteria (RAxML options "-N autoMR" [recommended], "-N autoMRE", and "-N autoMRE_IGN") and the frequency-based criterion (option "-N autoFC"). If you check the box **"BS brL"** (option "-k"), branch lengths will be saved in the bootstrap trees (which increases computation time). Additionally you can select the **number of independent ML searches** in the second and the fourth option.

You can load a Newick file to provide a **starting tree** for the ML-search through the analysis menu (RAxML option "-t").

With a nucleotide alignment you can load a file in which brackets ((), [], {}, <>) define stems and pseudo-knot regions of the sequence's **secondary structure**. The absolute number of characters of this file must be identical to the number of sites of the alignment. Nucleotide positions within the specified regions are represented by dots (".").

Additional Analyses

With a file containing a set of trees (e. g. the "RAxML_bootstrap" output file from a bootstrap analysis) you can generate majority rule or strict **consensus trees** (RAxML option "-J"). Also you can create a **set of pruned trees**, which lack a set of rogue taxa, that decrease support values (Pattengale et al. 2011), a consensus tree from these pruned trees is automatically created.

In addition you can compute **Robinson-Foulds pairwise distances** (Robinson & Foulds 1981) between trees (option "-f r").

With an alignment loaded and a file containing one or more trees you can also compute **per site log Likelihoods** (option "-f g"). The output that can be read by CONSEL[10] (Shimodaira 2001) to calculate p-values. In addition you can compute **SH-like support values** (Shimodaira & Hasegawa 1999) on a given (best) tree (option "-f J").

Excluding sites

In order to exclude sites you can use the interactive panel, or load a file which has to be in the RAxML standard format:

```
1-200 333-333 500-667
```

Loading a file with this content, or specifying these ranges in the panel, will produce a new alignment file (using the RAxML flag "-E"), reduced by the first 200 columns, the single site 333 and the sites 500–667. The new alignment file is automatically set as input and is again checked for identical sequences and/or gap-only characters.

Partitioning the matrix

There are three different ways of defining partitions (RAxML option "-q"):

Maybe the easiest way is to load individual partitions one after the other as separate files with the "Add Alignment" -button. RaxmlGUI will automatically set the partitions according to the determined data type. **CAUTION! RaxmlGUI will ONLY check for identical taxon counts! Make sure taxon order and names are identical in all files!**

The second way is to load the combined data set as one and to set partitions in an interactive way with the option "Set/Edit partitions..." .

[10] You can use the CONSEL command "seqmt --puzzle YOURFILE" to convert it to CONSEL format

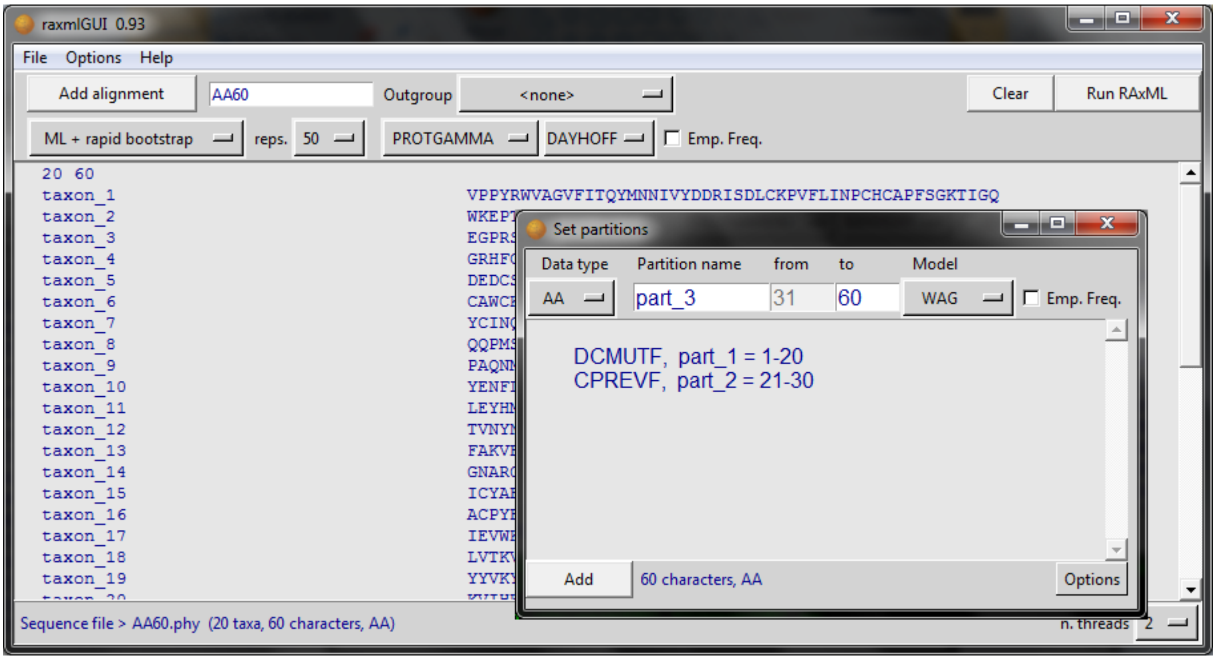

Figure 2: The partitioning panel

The third way is to load a file with the format specified in Stamatakis (2008)[11]; note that in a partitioned analysis every site has to be assigned to one partition, i. e. sites must not be assigned to two different partitions, and no site is allowed to be not assigned to any partition.

In any case you can edit, delete, and export the partitions set.

If you successfully set a partitioning scheme on your data set, you will be asked before the run starts if you want to calculate the branch lengths independently for each partition (RAxML option "`-M`"). This will produce best-scoring likelihood trees with branch length optimized for each partition (identical topology).

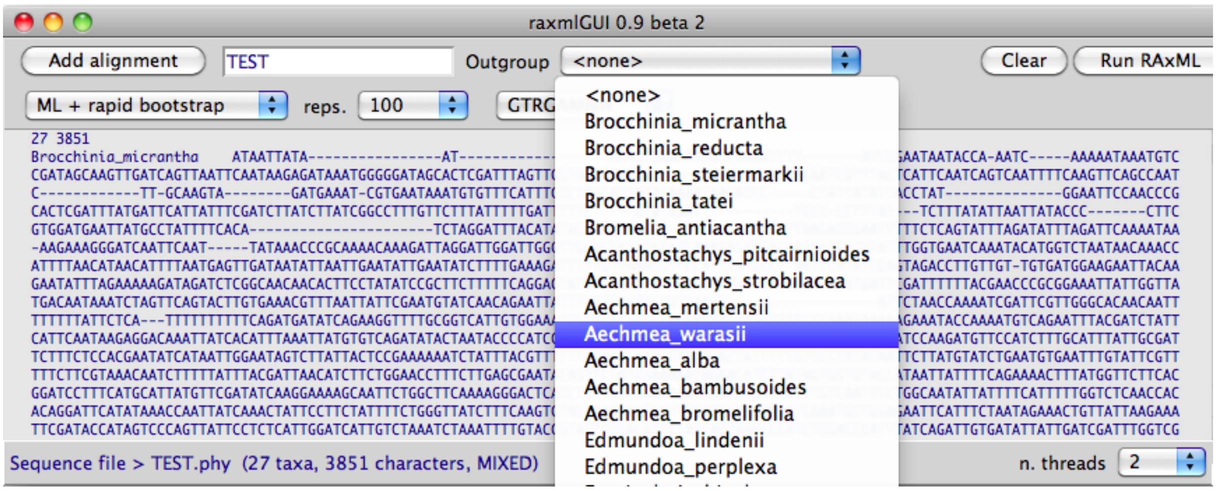

Figure 3: Outgroup selection

[11]See also Appendix B

Outgroup selection

A list of all taxa in the matrix is created in the program's toolbar for a quick single-taxon outgroup selection (RAxML option "`-o`"). If you want to specify more than one taxon to be in the outgroup, choose the option "select multiple outgroup" from the analysis menu. A window will open with a list of all taxa. You can mark more than one taxon by holding ⇧ (selecting ranges) or Ctrl (selecting single entries) for selecting those as outgroup. If RAxML finds the multiple outgroup not to be monophyletic, it will take the first taxon in the list as outgroup. If no outgroup is selected, the tree will be unrooted.

Setting a topological constraint

You can enforce topological constraints to your analysis through the menu option "enforce constraints". You can define taxon groups through a panel ("define topological constraint...") or by uploading a Newick formatted tree file, which can be binary or multifurcating. RAxML options "`-r`" or "`-g`" will be used for binary or multifurcating tree constraints, respectively.

Note that RAxML accepts only backbone constraints, which means, that unconstrained taxa can be placed at any position in the resulting tree, including within constrained clades. If you want to constrain monophyletic groups you can check the respective box in the panel, this will automatically append the remaining taxa to the set of constraints, so all constrained clades will result monophyletic.

Output files

You can set the name of the RAxML output files in the text field. By default the output file name is the same as the input file (without extension). The suffix "`_red`" is appended if identical sequences (or gap-only characters) are excluded from the analyses, and "`_exc`" is used when the "exclusion site" option is applied. If a RAxML info file with the same ID is found in the directory, you are prompted to change the output name before starting the analysis. For all types of output files, and their contents please refer to Stamatakis (2008).

Utilities

In some trees RAxML associates support values to the branches, rather than to the nodes (e. g. consensus trees, and those with SH-like support values). This format is not supported by FigTree (Rambaut 2006). However, if you want to use this program, you can produce a modified version of those tree files using the menu option **"Convert to FigTree format"**.

It is possible to **export your alignment in NEXUS format**. If the necessary python library DendroPy (Sukumaran & Holder 2010) is not yet installed, you will be guided through its installation.[6]

You can **inspect the RAxML command** that will be executed in the terminal before pressing "run RAxML" with the menu option "show RAxML command". In case of pipelined analyses this can contain many commands. Further it is possible to save the command(s) to a file.

You can **export the citation** for raxmlGUI in the following formats: Text, BiBTeX, EndNote (xml), and Reference manager (RIS).

Keyboard shortcuts

Action	Short cut Win/Linux	Short cut Mac
Open alignment	Ctrl O	⌘ O
Export alignment	Ctrl S	⌘ S
Change analysis type	Ctrl A	⌘ A
Import FASTA file	Ctrl F	⌘ F
Exclude sites	Ctrl E	⌘ E
Set partitions	Ctrl P	⌘ P
Clear	Ctrl ⇧ K	⇧ ⌘ K
Quit raxmlGUI	Ctrl Q	⌘ Q
Preferences	Ctrl ⇧ P	⌘ ,
Close window	Ctrl W	⌘ W
Run analysis	Ctrl R	⌘ R
Open this raxmlGUI help	Ctrl H	⌘ ?
Import NEXUS file	Ctrl N	⌘ N
Save changes (exclude sites/ define partitions/preferences)	↵	

Contacts

If you find any problems/bugs or want to give us a feedback, please contact us: raxmlgui.help@gmail.com.
If you want to get the latest news (e. g. new releases and updates), you can subscribe to our mailing list on http://lists.sourceforge.net/lists/listinfo/raxmlgui-news.
If you have problems or questions regarding RAxML, please have a look at the RAxML google group.

References

Berger, S. A. and Stamatakis, A. 2010. Accuracy and performance of single versus double precision arithmetics for maximum likelihood phylogeny reconstruction. *Parallel Processing and Applied Mathematics*. Ed. by R. Wyrzykowski, J. Dongarra, K. Karczewski, and J. Wasniewski. Vol. 6068. Lecture Notes in Computer Science, 270–279. doi: 10.1007/978-3-642-14403-5_29.

Huson, D. H., Richter, D. C., Rausch, C., Dezulian, T., Franz, M., and Rupp, R. 2007. Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics* **8**, 460. doi: 10.1186/1471-2105-8-460.

Maddison, D. R., Swofford, D. L., and Maddison, W. P. 1997. NEXUS: An Extensible File Format for Systematic Information. *Systematic Biology* **46**, 590–621. doi: 10.1093/sysbio/46.4.590.

Ott, M., Zola, J., Stamatakis, A., and Aluru, S. 2007. Large-scale maximum likelihood-based phylogenetic analysis on the IBM BlueGene/L. *Proceedings of the ACM/IEEE Supercomputing Conference*. Los Alamitos, CA, USA: IEEE Computer Society. doi: 10.1145/1362622.1362628.

Page, R. D. M. 1996. Tree View: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**, 357–358. doi: 10.1093/bioinformatics/12.4.357.

Pattengale, N. D., Aberer, A. J., Swenson, K. M., Stamatakis, A., and Moret, B. M. E. 2011. Uncovering hidden phylogenetic consensus in large datasets. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **8**, 902–911. doi: 10.1109/TCBB.2011.28.

Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R. P., Moret, B. M. E., and Stamatakis, A. 2010. How many bootstrap replicates are necessary? *Journal of Computational Biology* **17**, 337–354. doi: 10.1089/cmb.2009.0179.

Price, M. N., Dehal, P. S., and Arkin, A. P. 2010. FastTree 2 – Approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490. doi: 10.1371/journal.pone.0009490.

Rambaut, A. 2006. FigTree. url: http://tree.bio.ed.ac.uk/software/figtree/.

Robinson, D. R. and Foulds, L. R. 1981. Comparison of phylogenetic trees. *Mathematical Biosciences* **53**, 131–147. doi: 10.1016/0025-5564(81)90043-2.

Shimodaira, H. 2001. Multiple comparisons of log-likelihoods and combining nonnested models with applications to phylogenetic tree selection. *Communications in Statistics, Part A – Theory and Methods* **30**, 1751–1772. doi: 10.1081/STA-100105696.

Shimodaira, H. and Hasegawa, M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* **16**, 1114–1116.

Stamatakis, A. 2006a. Phylogenetic models of rate heterogeneity: a high performance computing perspective. *Proceedings of the 20th IEEE/ACM International Parallel and Distributed Processing Symposium (IPDPS2006), High Performance Computational Biology Workshop*. Rhodos, Greece. doi: 10.1109/IPDPS.2006.1639535.

Stamatakis, A. 2006b. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690. doi: 10.1093/bioinformatics/btl446.

Stamatakis, A. 2008. The RAxML 7.0.4 manual. Available at http://sco.h-its.org/exelixis/.

Stamatakis, A., Blagojevic, F., Nikolopoulos, D. S., and 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., and Rougemont, J. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology* **57**, 758–771. doi: 10.1080/10635150802429642.

Sukumaran, J. and Holder, M. T. 2010. DendroPy: a Python library for phylogenetic computing. *Bioinformatics* **26**, 1569–1571. doi: 10.1093/bioinformatics/btq228.

Yang, Z. 1994. Maximum Likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution* **39**, 306–314. doi: 10.1007/BF00160154.

python™

Appendix

A. A working PHYLIP example

```
6 40
1_first_row_reads_nroftaxa_blank_nrofcharacters GTGGCGGTCATTCTCATTTG
2_this_is_a_working_example__taxon_names_can_be_very_long ATTCGTGGTCATTCGTGGTC
3_not_allowed_in_names_are_blanks_and_special_characters CGTGACATTCGTGTCTTGGT
4_use_blanks_or_tabs_to_delimit_taxon_names_from_characters TCATTCGTGCGATGTCTGTG
5_there_is_no_option_for_comments_like_in_Nexus_format TGTTGCGTTGGTCATCTCAG
6_for_interleaved_format_dont_repeat_the_taxon_names TTTCTTGGGCGGTCGTTCAA

TCTCATTGGCGGTCATTGGT
ATGTTGGCGGTCATTCTCTG
ATTTGGCCTCCAGGTGTGTT
GGGTCACTCATTCGTGTTGT
TGTGATTGCGGTCATTCTCG
GTTTGGACTCATTTGCGGTC
```

B. The format of the Partition file

General format

Parttype, partname = partrange
e. g.:

```
DNA, partition1 = 1-100
BIN, partition2 = 101-200
```

A slightly more complex example

```
JTT, AAgene1 = 1-500

WAGF, AAgene2withempiricalfrequencies = 501-800

MULTI, multistatepartition3 = 801-900

BIN, binarypartition4 = 901-1000

DNA, DNAgene5codon1and2 = 1001-1500\3, 1002-1500\3
DNA, DNAgene5codon3 = 1003-1500\3

DNA, DNAgene6codon1 = 1501-2000\3
DNA, DNAgene6codon2 = 1502-2000\3
DNA, DNAgene6codon3 = 1503-2000\3

DNA, DNAsplitgene7 = 2001-2200, 2800-3000
DNA, intronofgene7 = 2201-2799
```

C. Selected papers that cite raxmlGUI

Antunes Carvalho, F. and Renner, S. S. 2012. A dated phylogeny of the papaya family (Caricaceae) reveals the crop's closest relatives and the family's biogeographic history. *Molecular Phylogenetics and Evolution* **65**, 46–53. doi: 10.1016/j.ympev.2012.05.019.

Blair, C., Davy, C. M., Ngo, A., Orlov, N. L., Shi, H-t., Lu, S-q., Gao, L., Rao, D-q., and Murphy, R. W. 2012. Genealogy and demographic history of a widespread amphibian throughout Indochina. *Journal of Heredity*. doi: 10.1093/jhered/ess079.

Chacón, J., Sousa, A., Baeza, C. M., and Renner, S. S. 2012. Ribosomal DNA distribution and a genus-wide phylogeny reveal patterns of chromosomal evolution in *Alstroemeria* (Alstroemeriaceae). *American Journal of Botany* **99**, 1501–1512. doi: 10.3732/ajb.1200104.

Deitz, K. C., Athrey, G., Reddy, M. R., Overgaard, H. J., Matias, A., Jawara, M., Della Torre, A., Petrarca, V., Pinto, J., Kiszewski, A. E., Kengne, P., Costanini, C., Caccone, A., and Slotman, M. A. 2012. Genetic isolation within the malaria mosquito *Anopheles melas*. *Molecular Ecology* **21**, 4498–4513. doi: 10.1111/j.1365-294X.2012.05724.x.

Desirò, A., Naumann, M., Epis, S., Novero, M., Bandi, C., Genre, A., and Bonfante, P. 2012. *Mollicutes*-related endobacteria thrive inside liverwort-associated arbuscular mycorrhizal fungi. *Environmental Microbiology*. doi: 10.1111/j.1462-2920.2012.02833.x.

Fernandes, A. M., Gonzalez, J., Wink, M., and Aleixo, A. 2012. Multilocus phylogeography of the Wedge-billed Woodcreeper *Glyphorynchus spirurus* (Aves, Furnariidae) in lowland Amazonia: widespread cryptic diversity and paraphyly reveal a complex diversification pattern. *Molecular Phylogenetics and Evolution* **66**, 270–282. doi: 10.1016/j.ympev.2012.09.033.

Fritz, U., Alcalde, L., Vargas-Ramírez, M., Goode, E. V., Fabius-Turoblin, D. U., and Praschag, P. 2012. Northern genetic richness and southern purity, but just one species in the *Chelonoidis chilensis* complex. *Zoologica Scripta* **41**, 220–232. doi: 10.1111/j.1463-6409.2012.00533.x.

Fritz, U., Corti, C., and Päckert, M. 2012. Mitochondrial DNA sequences suggest unexpected phylogenetic position of Corso-Sardinian grass snakes (*Natrix cetti*) and do not support their species status, with notes on phylogeography and subspecies delineation of grass snakes. *Organisms Diversity & Evolution* **12**, 71–80. doi: 10.1007/s13127-011-0069-8.

Fritz, U., Vargas-Ramírez, M., and Široký, P. 2012. Phylogenetic position of *Pelusios williamsi* and a critique of current GenBank procedures (Reptilia: Testudines: Pelomedusidae). *Amphibia-Reptilia* **33**, 150–154.

Hawlitschek, O. and Glaw, F. 2012. The complex colonization history of nocturnal geckos (*Paroedura*) in the Comoros Archipelago. *Zoologica Scripta*. doi: 10.1111/zsc.12001.

Hawlitschek, O., Hendrich, L., Espeland, M., Toussaint, E. F. A., Genner, M. J., and Balke, M. 2012. Pleistocene climate change promoted rapid diversification of aquatic invertebrates in Southeast Australia. *BMC Evolutionary Biology* **12**, 142. doi: 10.1186/1471-2148-12-142.

Hawlitschek, O., Nagy, Z. T., and Glaw, F. 2012. Island evolution and systematic revision of comoran snakes: why and when subspecies still make sense. *PLoS ONE* **7**, e42970. doi: 10.1371/journal.pone.0042970.

Jaklitsch, W. M. and Voglmayr, H. 2012. *Hypocrea britdaniae* and *H. foliicola*: two remarkable new European species. *Mycologia* **104**, 1213–1221. doi: 10.3852/11-429.

Koch, A., Huelsken, T., and Hoffmann, J. 2012. The Young Systematists special issue — promoting the scientific work of early career scientists in taxonomy and systematics. *Organisms Diversity & Evolution* **12**, 333–334. doi: 10.1007/s13127-012-0114-2.

Lavoué, S. 2012. *Petrocephalus* Marcusen, 1854 (Osteoglossomorpha: Mormyridae) of the Bangweulu-Mweru ecoregion (Luapula River system, Congo basin), with the description of a new species. *Journal of Natural History* **46**, 2159–2178. doi: 10.1080/00222933.2012.708449.

Lavoué, S., Miya, M., Arnegard, M. E., Sullivan, J. P., Hopkins, C. D., and Nishida, M. 2012. Comparable ages for the independent origins of electrogenesis in African and South American weakly electric fishes. *PLoS ONE* **7**, e36287. doi: 10.1371/journal.pone.0036287.

Licht, M., Schmücker, K., Huelsken, T., Hanel, R., Bartsch, P., and Päckert, M. 2012. Contribution to the molecular phylogenetic analysis of extant holocephalan fishes (Holocephali, Chimaeriformes). *Organisms Diversity & Evolution* **12**, 421–432. doi: 10.1007/s13127-011-0071-1.

Lutz, M., Vánky, K., and Piątek, M. 2012. *Shivasia* gen. nov. for the Australasian smut *Ustilago solida* that historically shifted through five different genera. *IMA Fungus* **3**, 143–154. doi: 10.5598/imafungus.2012.03.02.06.

Norlinder, E., Nygren, A., Wiklund, H., and Pleijel, F. 2012. Phylogeny of scale-worms (Aphroditiformia, Annelida), assessed from 18SrRNA, 28SrRNA, 16SrRNA, mitochondrial cytochrome c oxidase subunit I (COI), and morphology. *Molecular Phylogenetics and Evolution* **65**, 490–500. doi: 10.1016/j.ympev.2012.07.002.

Päckert, M., Martens, J., Liang, W., Hsu, Y-C., and Sun, Y-H. 2012. Molecular genetic and bioacoustic differentiation of *Pnoepyga* Wren-babblers. *Journal of Ornithology*. doi: 10.1007/s10336-012-0897-0.

Pérez-Ortega, S., Fernández-Mendoza, F., Raggio, J., Vivas, M., Ascaso, C., Sancho, L. G., Printzen, C., and Ríos, A. de los 2012. Extreme phenotypic variation in *Cetraria aculeata* (lichenized Ascomycota): adaptation or incidental modification? *Annals of Botany* **109**, 1133–1148. doi: 10.1093/aob/mcs042.

Pessoa, E. M., Alves, M., Alves-Araújo, A., Palma-Silva, C., and Pinheiro, F. 2012. Integrating different tools to disentangle species complexes: a case study in *Epidendrum* (Orchidaceae). *Taxon* **61**, 721–734.

Poczai, P., Hyvönen, J., Taller, J., Jahnke, G., and Kocsis, L. 2012. Phylogenetic analyses of Teleki Grapevine rootstocks using three chloroplast DNA markers. *Plant Molecular Biology Reporter*. doi: 10.1007/s11105-012-0512-9.

Printzen, C., Fernández-Mendoza, F., Muggia, L., Berg, G., and Grube, M. 2012. Alphaproteobacterial communities in geographically distant populations of the lichen *Cetraria aculeata*. *FEMS Microbiology Ecology* **82**, 316–325. doi: 10.1111/j.1574-6941.2012.01358.x.

Terry, R. G., Bartel, J. A., and Adams, R. P. 2012. Phylogenetic relationships among the New World cypresses (*Hesperocyparis*; Cupressaceae): evidence from noncoding chloroplast DNA sequences. *Plant Systematics and Evolution* **10**, 1987–2000. doi: 10.1007/s00606-012-0696-3.

Ure, A. E. and Forslund, O. 2012. Lack of methylation in the upstream region of human papillomavirus type 6 from aerodigestive tract papillomas. *Journal of Virology* **86**, 13790–13794. doi: 10.1128/JVI.01938-12.

Vargas-Ramírez, M., Michels, J., Castaño-Mora, O. V., Cárdenas-Arevalo, G., Gallego-García, N., and Fritz, U. 2012. Weak genetic divergence between the two South American toad-headed turtles *Mesoclemmys dahli* and *M. zuliae* (Testudines: Pleurodira: Chelidae). *Amphibia-Reptilia* **33**, 373–385. doi: 10.1163/15685381-00002840.

Velo-Antón, G., Godinho, R., Harris, D. J., Santos, X., Martínez-Freiria, F., Fahd, S., Larbes, S., Pleguezuelos, J. M., and Brito, J. C. 2012. Deep evolutionary lineages in a Western Mediterranean snake (*Vipera latastei*/*monticola* group) and high genetic structuring in Southern Iberian populations. *Molecular Phylogenetics and Evolution* **65**, 965–973. doi: 10.1016/j.ympev.2012.08.016.

Voglmayr, H., Rossman, A. Y., Castlebury, L. A., and Jaklitsch, W. M. 2012. Multigene phylogeny and taxonomy of the genus *Melanconiella* (Diaporthales). *Fungal Diversity* **57**, 1–44. doi: 10.1007/s13225-012-0175-8.

Wagner, N., Silvestro, D., Brie, D., Ibisch, P. I., Zizka, G., Weising, K., and Schulte, K. 2012. Spatiotemporal evolution of *Fosterella* (Bromeliaceae) in the Central Andean biodiversity hotspot. *Journal of Biogeography*. doi: 10.1111/jbi.12052.

Yurkov, A. M. and Golubev, W. I. 2012. Phylogenetic study of *Cryptococcus laurentii* mycocinogenic strains. *Mycological Progress*. doi: 10.1007/s11557-012-0873-3.

D. References for Amino Acid substitution models

DAYHOFF: Dayhoff, M. O., Schwartz, R. M., and Orcutt, B. C. 1978. A model of evolutionary change in proteins. *Atlas of Protein Sequence and Structure*. Ed. by M. O. Dayhoff. Vol. 5 suppl. 2. National Biomedical Research Foundation, 345–352.

DCMUT: Kosiol, C. and Goldman, N. 2005. Different versions of the Dayhoff rate matrix. *Molecular Biology and Evolution* **22**, 193–199. doi: 10.1093/molbev/msi005.

JTT: Jones, D. T., Taylor, W. R., and Thornton, J. M. 1992. The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences* **8**, 275–282. doi: 10.1093/bioinformatics/8.3.275.

MTREV: Adachi, J. and Hasegawa, M. 1995. Model of amino acid substitution in proteins encoded by mitochondrial DNA. *Journal of Molecular Evolution* **42**, 459–468. doi: 10.1007/BF02498640.

WAG: Whelan, S. and Goldman, N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution* **18**, 691–699.

RTREV: Dimmic, M. W., S., Rest J., P., Mindell D., and A., Goldstein R. 2002. rtREV: a substitution matrix for inference of retrovirus and reverse transcriptase phylogeny. *Journal of Molecular Evolution* **55**, 65–73. doi: 10.1007/s00239-001-2304-y.

CPREV: Adachi, J., Waddell, P., Martin, W., and Hasegawa, M. 2000. Plastid genome phylogeny and a model of amino acid substitution for proteins encoded by chloroplast DNA. *Journal of Molecular Evolution* **50**, 348–358. doi: 10.1007/s002399910038.

VT: Müller, T. and Vingron, M. 2000. Modeling amino acid replacement. *Journal of Computational Biology* **7**, 761–776. doi: 10.1089/10665270050514918.

BLOSUM62: Henikoff, S. and Henikoff, J. G. 1992. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the U.S.A.* **89**, 10915–10919.

MTMAM: Yang, Z., Nielsen, R., and Hasegawa, M. 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. *Molecular Biology and Evolution* **15**, 1600–1611.

LG: Le, S. and Gascuel, O. 2008. An improved general amino acid replacement matrix. *Molecular Biology and Evolution* **25**, 1307–1320. doi: 10.1093/molbev/msn067.

MTART: Abascal, F., Posada, D., and Zardoya, R. 2007. MtArt: a new model of amino acid replacement for Arthropoda. *Molecular Biology and Evolution* **24**, 1–5. doi: 10.1093/molbev/msl136.

MTZOA: Rota-Stabelli, O., Yang, Z., and Telford, M. J. MtZoa: a general mitochondrial amino acid substitutions model for animal evolutionary studies. *Molecular Phylogenetics and Evolution* **52**, 268–272. doi: 10.1016/j.ympev.2009.01.011.

PMB: Veerassamy, S., Smith, A., and Tillier, E. R. 2003. A transition probability model for amino acid substitutions from blocks. *Journal of Computational Biology* **10**, 997–1010. doi: 10.1089/106652703322756195.

HIVB, HIVW: Nickle, D. C., Heath, L., Jensen, M. A., Gilbert, P. B., Mullins, J. I., and Kosakovsky Pond, S. L. 2007. HIV-specific probabilistic models of protein evolution. *PLoS ONE* **2**, e503. doi: 10.1371/journal.pone.0000503.

FLU: Dang, C. C., Le, Q. S., Gascuel, O., and Le, V. S. 2010. FLU, an amino acid substitution model for influenza proteins. *BMC Evolutionary Biology* **10**, 99.
doi: 10.1186/1471-2148-10-99.

GTR: Tavaré, S. 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences* **17**, 57–86.

LaTeX