

## Program note: YBYRÁ, a fast and resourceful tool for sensitivity analysis

**Machado, Denis Jacob<sup>1</sup>**

<sup>1</sup>University of São Paulo, Institute of Biosciences, Inter-institutional Grad Program on Bioinformatics. Email: [machadodj@usp.br](mailto:machadodj@usp.br).

**Abstract:** YBYRÁ is a general and useful application to evaluate the sensitivity of tree topology to different tree search strategies, optimality criteria, alignment methods, transformation cost schemes, or any other variable that may affect hypothesis choice. It provides tools for automatic generation of sensitivity plots (“Navajo rugs”) and tree distance matrices based on shared splits/ clades. It also renders a framework to facilitate the search for potential rogue taxa based on how much they influence average matching split distances.

**Key words:** phylogenetics, sensitivity analysis, tree distance, wildcard.

In phylogenetic systematics, authors make use sensitivity analysis to address how much hypothesis choice may be affected by variables such as different tree search strategies, optimality criteria, alignment methods, and transformation cost schemes (e.g., Higdon et al. 2007; Miller et al. 2010 ; Payne 2014). There is some debate in the literature regarding the scientific and heuristic value of sensitivity analysis (Grant & Kluge 2005; Giribet & Wheeler 2007). However, the instrumental value of sensitivity analysis as means to describe and compare different methodological approaches in systematics is indisputable.

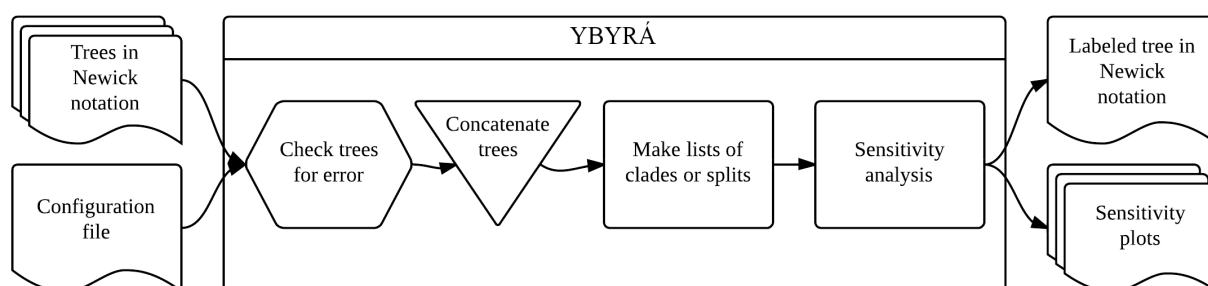
The program “YBYRÁ” was written to facilitate comparison of tree topologies obtained using different methods and parameters. Although there are already programs dedicated to sensitivity analysis in systematics (such as “Cladescan”, Sanders 2010), the reader may find YBYRÁ useful due to its velocity and resourcefulness.

### Program description

YBYRÁ is written in Python; hence, it is a cross-platform application (e.g., Windows, MAC or Linux) and does not require compilation. Although it can be executed as a stand-alone application, YBYRÁ optionally makes use of free, easy to install Python modules to root trees and print sensitivity plots (“Navajo rugs”). The program, detailed configuration instructions, sample input files, and tutorial exercises using most of the available options can be down-

loaded for free at <http://goo.gl/9C5kq>. This provisional URL will soon be replaced by a permanent URL where YBYRÁ will be distributed for free under the GNU General Public Licence.

The input to YBYRÁ consists of (1) one or more tree files, (2) one configuration file, and (3) optional command line arguments. The program accepts rooted or unrooted dichotomous trees, as well as trees with different roots and/ or terminals. Operationally, lists of clades or splits are obtained from the provided trees. Those lists are compared to each other and the results are given in a variety of ways, including sensitivity plots (“Navajo rugs”), spreadsheets and annotated trees with branch lengths representing the frequency of a particular clade or split (see [Fig. A1](#)).



**Figure A1:** Simple implementation of sensitivity analysis in YBYRÁ. Input consists of a configuration file and one or more files containing trees in Newick notation. YBYRÁ will check for errors in the trees and make a safe copy of them into a new file. According to user options, clades or splits will be stored in the form of lists that will be used for the sensitivity analysis. The main output consists of a labeled tree in Newick notation and sensitivity plots in SVG format. Image files will be given names according to node numbers to facilitate editing.

YBYRÁ was written to allow sensitivity analysis in large data sets. Sensitivity analysis in YBYRÁ can be executed faster than in Cladescan and with fewer memory requirements, facilitating analysis of large data sets (see [Table A1](#)).

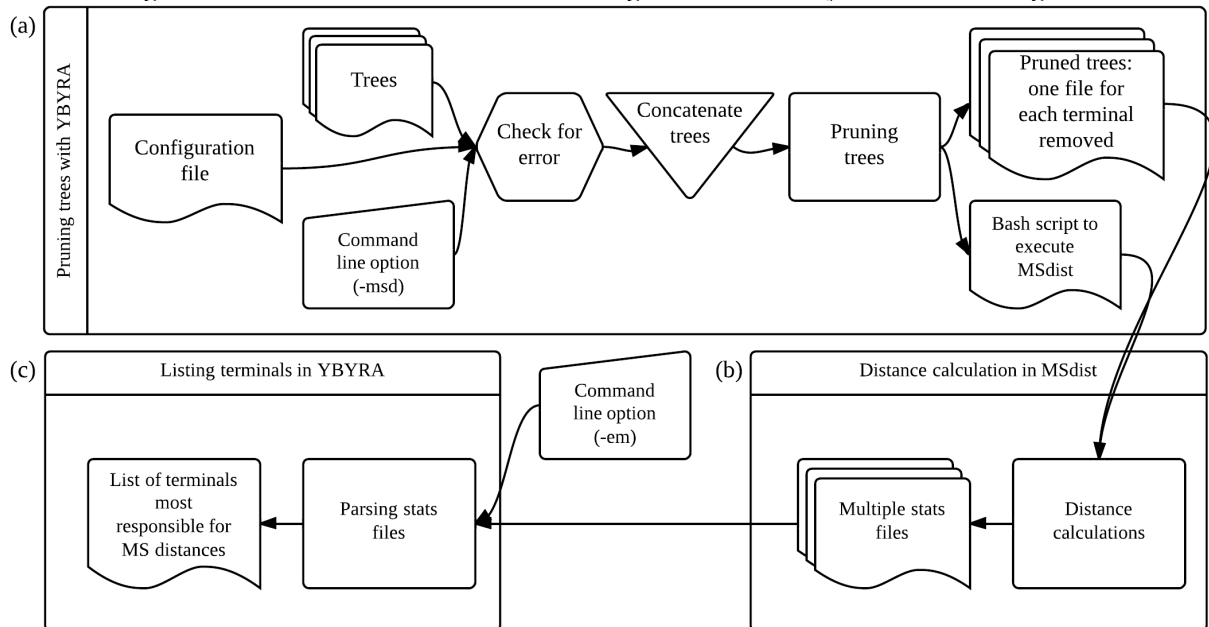
**Table A1:** Comparing the approximate execution time in seconds for sensitivity analysis in Cladescan and YBYRÁ. All clades in the trees were examined by each program. All analysis were performed in a 2.9 GHz Intel Core i7 with 8 GB (1600 MHz DDR3) of memory.

No. of terminals	No. of trees	Approx. execution time (sec.)	
		Cladescan	YBYRÁ
10	1,000	4	3
100	100	13	3
1,000	10	>9h	11

YBYRÁ also offers options to facilitate the search for rogue taxa according to how much they affect the average match split distances (Bogdanowicz & Giaro 2012) calculated in

MSdist 0.5 (<http://www.kaims.pl/~dambo/msdist/>). First, trees are pruned, one terminal at a time, generating concatenated tree files (**Fig. A2.a**). Second, pruned trees are subjected to MSdist to calculate the average matching split distances for each terminal that was removed (**Fig. A2.b**). Finally, YBYRÁ uses MSdist outputs to generate an ordered list of terminals according to how much they affect the average matching split distances (**Fig. A2.c**). Terminals that most affect match split distances (i.e., terminals that result in the lowest average match split distance when removed) are more likely to cause resolution decrease in the strict consensus trees and should be examined as potential wildcard.

**Figure A2:** Example implementation of rogue taxa search using YBYRÁ and MSdist 0.5: (a) YBYRÁ will make make copies of the original trees removing one terminal at a time. YBYRÁ will also print a bash script to execute MSdist; (b) average match split (MS) distances are calculated in MSdist; (c) MSdist results are used by YBYRÁ to generate an ordered list of terminals according to how much they influence average MS distances.



Other options are available, including distance calculation based on shared splits/clades between pairs of trees. An integral list of options with default values and dependencies is presented in **Table A2**. To facilitate usage, YBYRÁ accompanies the “makeYbyraConfiguration” program to create and edit configuration files. The makeYbyraConfiguration is a cross-platform graphic user interface written in Python using the TKinter module and can be executed directly from the command line.

**Table A2:** Complete list of YBYRÁ's options with default values and dependencies.

<b>Manifest options</b>	<b>* Expects</b>	<b>Default</b>	<b>Dependencies</b>
>ID=*	Text	output	None
>Root=*	Text	EMPTY	Biopython
<Begin File * End Files>	File names separated by semicolons	EMPTY	None
<Begin Clades * End Clades>	Lists of terminal names. Terminal names separated by “,” and lists by “;”	EMPTY	None
<Begin Trees* End Trees>	Trees (Newick format) separated by “;”	EMPTY	None
>N=*	Integer	1	None
<Begin Pools * End Pools>	Lists of trees. Tree numbers separated by “;”, use “-” for ranges. Lists separated by “;”	EMPTY	None
>Opt=*	Integer 1) Look for topologies equal to Tree No. N 2) Lists every clade or split on each tree 3) Compare trees measuring the number of splits they share with each other 4) Compare trees calculating the number of splits they share with all the list of splits that are in every tree	1	None
>Compare=*	0 or 1 0) Do sensitivity analysis using splits 1) Do sensitivity analysis using clades (root dependent)	0	None
>SA=*	Yes or No. Sensitivity analysis on/ off	No	None
>RUG=*	Yes or No. Sensitivity plots on/ off	No	svgwrite
>Stroke=*	Stroke color as text or rgb code	Black	svgwrite
>Color=*	Background color as text or rgb code	Black	svgwrite
>Text=*	P or R. Print sensitivity plots with percentages (P) or ratios (R)	P	svgwrite
>Verbose*	Prints additional information on the screen while executing YBYRÁ	EMPTY	None
<b>Argument options</b>	<b>Description</b>	<b>Default</b>	<b>Dependencies</b>
-help (or -h)	Displays program help	DISABLE	None
-different_terminals (or -dt)	Trim terminals that are not in every tree	DISABLE	Biopython
-msdist (or -msd)	Create pruned trees to search for taxa that most affect match split distances using MSDist	DISABLE	Biopython
-examine_msdist (or -em)	Writes the list terminals that most affect match split distances using MSdis outputs	DISABLE	None

## Acknowledgements

YBYRÁ was first published as a poster at the XXXII Meeting of the Willi Hennig Society (Rostock, Germany, 2013) with financial support from FAPESP (Proc. No. 2013/05958-8). I thank Fernando Portella de Luna Marques and Taran Grant for their insightful suggestions. Ybyrá is a noun in Tupi which means wood, stick, tree, rod, stalk, lance or spear. I thank Miguel Trefault Rodrigues for suggesting this name.

## References

- Bogdanowicz, D., & Giaro, K. (2012). Matching split distance for unrooted binary phylogenetic trees. *IEEE/ACM Transactions on Computational Biology and Bioinformatics / IEEE, ACM*, 9(1), 150–60.
- Giribet, G., & Wheeler, W. C. (2007). The case for sensitivity: a response to Grant and Kluge. *Cladistics*, 23(3), 294–296.
- Grant, T., & Kluge, A. (2005). Stability, sensitivity, science and heurism. *Cladistics*, 21, 597–604.
- Higdon, J. W., Bininda-Emonds, O. R. P., Beck, R. M. D., & Ferguson, S. H. (2007). Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multi-gene dataset. *BMC Evolutionary Biology*, 7, 216.
- Miller, J. a, Carmichael, A., Ramírez, M. J., Spagna, J. C., Haddad, C. R., Rezác, M., ... Griswold, C. E. (2010). Phylogeny of entelegyne spiders: affinities of the family Penestomidae (NEW RANK), generic phylogeny of Eresidae, and asymmetric rates of change in spinning organ evolution (Araneae, Araneoidea, Entelegynae). *Molecular Phylogenetics and Evolution*, 55(3), 786–804.
- Payne, A. (2014). Resolving the relationships of apid bees (Hymenoptera: Apidae) through a direct optimization sensitivity analysis of molecular, morphological, and behavioural characters. *Cladistics*, 30(1), 11–25.
- Sanders, J. G. (2010). Program note: Cladescan, a program for automated phylogenetic sensitivity analysis. *Cladistics*, 26(1), 114–116.