

**MONOPHYLER: A program for calculating monophyly probabilities of
sampled lineages given a species tree**

Rohan Mehta
February 28, 2022

CONTENTS

1. Introduction	2
2. Getting Started	2
2.1. Windows	3
2.2. Mac	3
2.3. Linux	3
3. Running MONOPHYLER	3
3.1. Starting the GUI	3
3.2. Command Line	3
4. Using the GUI	4
4.1. Input from file	4
5. Format of input file	5
5.1. Number of species	6
5.2. Tree topology	6
5.3. Class-S sample sizes	7
5.4. Class-C sample sizes	7
5.5. Branch lengths	7
5.6. Monophyly event	7
5.7. File input example, GUI	8
6. Manual Input	9
6.1. Calculation menu	11
6.2. Number of species box	11
6.3. Tree input box	11
6.4. S input box	11
6.5. Branch length box	12
6.6. C input box	13
6.7. Monophyly event menu	13
6.8. Tree, S , branch length, and C status boxes	13
6.9. Main status box	13
6.10. Send button	13

6.11. Clear button	13
6.12. Example button	13
7. Performing a manual input calculation	13
8. Joint Monophyly	16
9. Errors to Ignore	17
10. Troubleshooting	18
11. Updates	18
12. References	18

1. INTRODUCTION

MONOPHYLER is a software package that computes probabilities of monophyly of gene lineages given a species tree. The primary computation that MONOPHYLER performs is the probability that a particular subsample of gene lineages is monophyletic for a particular locus given a species tree. This probability, which is obtained using the coalescent model of the evolution of gene trees conditional on species trees, is eq. 7 of Mehta, Bryant, and Rosenberg (2016). For this computation, it is assumed that gene lineages are sampled from each species and that the user is interested in the event that a certain subset of those samples are monophyletic when a gene tree of all samples is constructed. The subset of sampled lineages for which the monophyly probability is desired is labeled S , and the remaining sampled lineages are labeled C . Eq. 7 of Mehta, Bryant, and Rosenberg (2016) can therefore be described as probability of monophyly of S lineages. MONOPHYLER can also compute eqs. 8 and 9 of Mehta, Bryant, and Rosenberg (2016), which calculate the probability of monophyly of C lineages and the probability of reciprocal monophyly of S and C lineages, respectively (for details, see Mehta, Bryant, and Rosenberg (2016)). Finally, MONOPHYLER can also compute the probability of reciprocal monophyly for all species for any set of gene samples from two (Rosenberg (2003) eq. 14), three (Rosenberg (2002) eq. 12, Mehta and Rosenberg (2019)), or four species (Mehta and Rosenberg (2019)). There is also a separate python script for computing joint monophyly probabilities (Mehta, Steel and Rosenberg (in prep.)).

2. GETTING STARTED

MONOPHYLER is available for Windows and Linux and is distributed fully-functional; no installation is necessary.

2.1. **Windows.** To use MONOPHYLER on a Windows computer, unzip the downloaded folder into the directory of your choice. Double-click on the file `monophyler_windows.exe`.

2.2. **Mac.** Currently not implemented due to unexpected C++ library incompatibilities.

2.3. **Linux.** To use MONOPHYLER on a Linux computer, unzip the downloaded folder into the directory of your choice. Make sure that the path of the chosen directory **does not contain a space in any folder name**. In a terminal, go to your chosen directory and run the command: `./monophyler_linux.sh` with additional arguments if necessary (see Section 3).

3. RUNNING MONOPHYLER

There are two ways to use MONOPHYLER: through a GUI and with the command line.

3.1. **Starting the GUI.** For Windows, double-clicking the executable file `monophyler_windows.exe` will start the GUI. You must use the command line if you do not want to use the GUI.

For Linux, run the shell script `monophyler_linux.sh` from a command line with no other arguments to start the GUI.

3.2. Command Line.

3.2.1. *Windows.* Open Command Prompt and navigate to the directory containing the MONOPHYLER files.

`start monophyler_windows.exe` will automatically start the GUI.

`start monophyler_windows.exe inputfilename` will use `inputfilename` as the name of the input file, and the output file will be called `output.txt`.

`start monophyler_windows.exe inputfilename outputfilename` will use `inputfilename` as the name of the input file and `outputfilename` as the name of the output file.

In all cases, the output file is overwritten when the program is run. **When running Monophyler from the command line in Windows, the program may not display progress and may allow the user to perform other commands with Command Prompt while Monophyler is still running. As such, the output file will appear incomplete until Monophyler is finished and, in addition, error messages may not appear. We recommend using the GUI on Windows.**

3.2.2. *Mac.* Currently not implemented due to unexpected C++ library incompatibilities.

Number	Code	Calculation	Source
1	sc	Monophyly of a subset or its complement Reciprocal monophyly of a single subset and its complement	Mehta, Bryant, and Rosenberg (2016)
2	2rm	Reciprocal monophyly of two species	Rosenberg (2003)
3	3rm	Reciprocal monophyly of three species	Rosenberg (2002), Mehta and Rosenberg (in prep.)
4	4rmb	Reciprocal monophyly of four species, balanced topology	Mehta and Rosenberg (in prep.)
5	4rmc	Reciprocal monophyly of four species, caterpillar topology	Mehta and Rosenberg (in prep.)

TABLE 1. The calculations performed by MONOPHYLER. Calculation 1 is called “subset-complement” here.

3.2.3. *Linux*. Open a terminal window and navigate to the directory containing the MONOPHYLER files.

`./monophlyer_linux.sh` will automatically start the GUI.

`./monophlyer_linux.sh inputfilename` will use `inputfilename` as the name of the input file, and the output file will be called `output.txt`.

`./monophlyer_linux.sh inputfilename outputfilename` will use `inputfilename` as the name of the input file and `outputfilename` as the name of the output file.

In all cases, the output file is overwritten when the program is run.

4. USING THE GUI

The main window of the GUI is used to perform calculations either listed in an input file or entered manually by the user.

4.1. **Input from file.** The main window of the GUI has text boxes for the user to type in the name of the input file (which contains a list of calculations to be performed; for the appropriate file format, see Section 5) and the name of the output file (which will list the results of the calculation). We recommend this approach if the user wishes to perform multiple calculations at one time.

Figure 3 shows the components of the main window of the program, including where the input and output files should go. Refer to this figure for the names of each component of the main window. The default input file name is `input.txt` and the default output file name is `output.txt`. The output file will be **semicolon-separated** and therefore easy to import into a spreadsheet editor such as Excel or data analysis software such as R. We recommend the user uses the extension `.txt` for the output file.

Once the input and output file names are provided, the user should press the **Run** button on the main window to start the calculations. The main status box and the progress bar

should indicate the progress of the calculations as well as any errors the program encounters when attempting to read the input file. If the program encounters an error, it will quit the calculations, but the results of all calculations before the error will still be written to the output file.

5. FORMAT OF INPUT FILE

It is very important that the input file is formatted properly. We provide example input files for each type of calculation that can be performed with this software; these files are called `[calculation]_example.txt` with the phrase in `[calculation]` being the calculation code (see Table 1).

The input file is divided into two parts: the header and the calculations (see Figure 1). Full-line comments (indicated by `//`) are allowed anywhere in the file except the first line. The first line of the input file (starting the header) must be:

```
Input file for Monophyler version 1.0 7/21/16
```

with no spaces at the end. **This line does not change with updates.**

Every line in the input file must be separated by a newline character (so typing “Enter” or “Return” after you finish a line is sufficient). The next line must be a positive integer indicating the total number of calculations present in the input file. Also, if this number is too large, the program will appear to be stuck when it reaches the end of the file but it in fact has completed the entire file. The line following this number must be an integer from 1 to 5 which indicates the calculation the user wishes to perform, detailed in Table 1. It is only possible to perform one type of calculation for any particular input file.

After the calculation indicator there must be a **non-commented** buffer line, which can be blank or any kind of text; the example files have the phrase `END HEADER`. This line marks the end of the header.

An example header looks like:

```
Input file for Monophyler version 1.0 7/21/16
100
1
END HEADER
```

This header indicates that 100 calculations are listed in this file and the calculations that will be performed involve subset-complement monophyly (as opposed to reciprocal monophyly of two, three, or four species).

Immediately following the header are the calculations. Each line describes a separate calculation. The calculations must contain the following information in order, with the different components separated by whitespace:

- Number of species in the species tree
- Topology of the species tree
- Class-S sample sizes
- Class-C sample sizes (only for subset-complement calculation)
- Branch lengths
- Monophyly event (only for subset-complement calculation)

An example calculation looks like:

```
3 "((AB)C)" 1,2,3 4,2,1 1,1.5,2,5001,3 SC
```

This calculation corresponds to Figure 2B.

5.1. Number of species. This number must be a positive integer and must match the number of leaves in the species tree given in the next argument. In the example calculation above this number is 3.

5.2. Tree topology. This input must be a string of parentheses and nonnumeric characters surrounded by a single pair of quotation marks. The leaves of the species tree are indicated by single nonnumeric characters and the topology of the species tree is determined by the location of parentheses. Each set of open and closed parentheses demarcates a subtree of the species tree (including the species tree itself). Leaves must not be surrounded by their own set of parentheses. Examples of correct tree topology inputs are given in Figure 2. In the example calculation above this tree is "((AB)C)".

An easy way to obtain the correct tree topology format is to start with a binary tree in Newick format (no multifurcation; you can artificially induce multifurcation by setting some branch lengths to zero when specifying branch lengths, but not here):

```
(Bovine:0.69395,(Hylobates:0.36079,(Pongo:0.33636,(G.Gorilla:0.17147,(P.paniscus:0.19268,H.sapiens:0.11927):0.08386):0.06124):0.15057):0.54939,Rodent:1.21460);
```

remove all branch length information (numbers and colons):

```
(Bovine,(Hylobates,(Pongo,(G.Gorilla,(P.paniscus,H.sapiens))))),Rodent);
```

remove all commas, semicolons, and whitespace:

```
(Bovine(Hylobates(Pongo(G.Gorilla(P.paniscusH.sapiens))))Rodent)
```

represent every species with a single character:

```
(B(H(O(G(PH))))R)
```

and finally put quotation marks around everything:

```
"(B(H(O(G(PH))))R)"
```

This example Newick tree was taken from <http://evolution.genetics.washington.edu/philip/newicktree.html>.

5.3. Class-S sample sizes. This input must be a string of positive integers separated by commas. These integers represent the numbers of lineages of class S in each leaf of the tree, with the leaves in the same order they were in the tree topology input (Section 5.2). In the example calculation above, these sample sizes are 1,2,3.

5.4. Class-C sample sizes. This input must be a string of positive integers separated by commas. These integers represent the numbers of lineages of class C in each leaf of the tree, with the leaves in the same order they were in the tree topology input (Section 5.2). This input is present **only** for the subset-complement calculation; adding it in for the other calculations will result in errors. In the example calculation above, these sample sizes are 4,2,1.

5.5. Branch lengths. This input is a string of positive real numbers separated by commas. These numbers represent the lengths of the branches of the tree given by the tree topology input (Section 5.2). The order of the branches corresponds to an in-order traversal of the species tree: if you take a diagram of the species tree oriented with the root at the top and the leaves at the bottom and move a vertical line across the diagram from left to right, every time the line touches a new node, you write down the length of the branch corresponding to that node. **Please use 5001 for an infinite branch length (generally just the root).** Examples of what this input would be for various different species trees are given in Figure 1. **Please be careful with this step; it is very easy to put internal branch lengths in the wrong places.**

In the example calculation above, the branch lengths are 1,1.5,2,5001,3.

5.6. Monophyly event. This input must be one of the following strings: S, C, or SC. These values correspond to the monophyly events (from Mehta, Bryant, and Rosenberg 2016) E_S , E_C , and E_{SC} , respectively, and determine whether you want class- S , class- C , or both classes monophyletic, respectively. This input is present **only** for the subset-complement calculation. Unlike with C sample sizes, the program will just ignore this input for the other calculations.

In the example calculation above, the monophyly condition is SC.

5.7. File input example, GUI. Figure 1 displays an example input file with four example calculations. This file (with comments) is distributed with this software as `sc_example.txt`. This file only covers the subset-complement calculation. Example files covering the other four calculations are also provided (see Figures 12-15). Please note that the third calculation in the subset-complement example file and the last calculations in the example files for three- and four-species reciprocal monophyly may take much longer than the others. Figure 2 displays the species trees and parameter settings that correspond to the four calculations for the subset-complement (sc) example file.

Notepad will sometimes display these input files strangely (it will remove line breaks, for instance). We recommend viewing and editing the example input files in another text editor. For Windows users, Notepad++ and WordPad work fine. Microsoft Word may also work. Notepad works perfectly well if you want to create your own input file from scratch.

```

Input file for Monophyler version 1.0 7/21/16
4
1
END HEADER
2 "(AB)" 1,2 3,1 1,5001,2 S
3 "((AB)C)" 1,2,3 4,2,1 1,1.5,2,5001,3 SC
5 "(((AB)(CD))E)" 0,1,3,0,2 2,0,1,3,4 0.2,1,0.3,2,0.5,1.2,0.6,5001,5 C
5 "((AB)(C(DE)))" 1,2,0,3,4 0,0,1,2,1 0.2,1,0.3,5001,0.5,2,0.6,1.2,5 S

```

FIGURE 1. Example input file for subset-complement calculation. Everything above and including the phrase `END HEADER` is the header and everything below is the set of calculations.

To perform these four calculations, run `MONOPHYLER` (see Section 3). Figure 3 displays the starting window of the program with components labeled. The input and output file name boxes are pre-filled with the default names (`input.txt` and `output.txt`, respectively). The default selection in the input method box is `input from file`. When this method for retrieving the input is selected, the input status box indicates that the program is ready to input from the given input file and to output the given output file. If either entry is blank, the input status box will ask the user to provide the appropriate file name and the program will not attempt to compute anything until both file names are provided. For this example, we change the input file name to `sc_example.txt` and the output file name to `sc_example_output.txt`. The changes are displayed in Figure 4.

To perform the calculations from the input file, press the `Run` button. The output is displayed in Figure 5. The main status box indicates which calculations are completed in

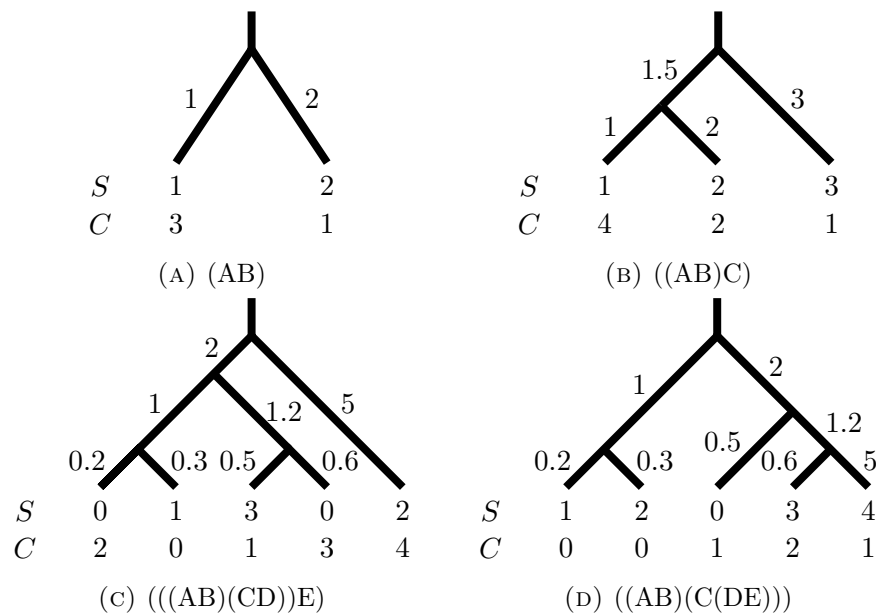


FIGURE 2. Species trees and parameters for calculations in `sc_example.txt` and Figure 1. Note that branch lengths are not to scale.

real time as well as how long it took to perform each calculation. The resulting output file is shown in Figure 6.

The output file lists the line number of the input file for a particular calculation, the type of the calculation (see Table 1). The output file indicates that the probability of monophyly of *S* lineages for the calculation presented in Figure 2a is 0.08183857, the probability of reciprocal monophyly for the calculation presented in Figure 2b is 1.45995×10^{-7} , the probability of monophyly of *C* lineages for the calculation presented in Figure 2c is 7.64864×10^{-9} , and the probability of monophyly of *S* lineages for the calculation presented in Figure 2d is 7.57277×10^{-8} . All of these probabilities are fairly low, which is to be expected due to the short branch lengths and non-species-specific sampling in Figure 2.

6. MANUAL INPUT

MONOPHYLER also provides the option of performing single calculations individually via user input into the GUI (not from a file). This feature can be accessed by choosing **Manual input** from the input method box in the main window of the program. This choice will open a new window where the user can put in the parameters for a single calculation. An

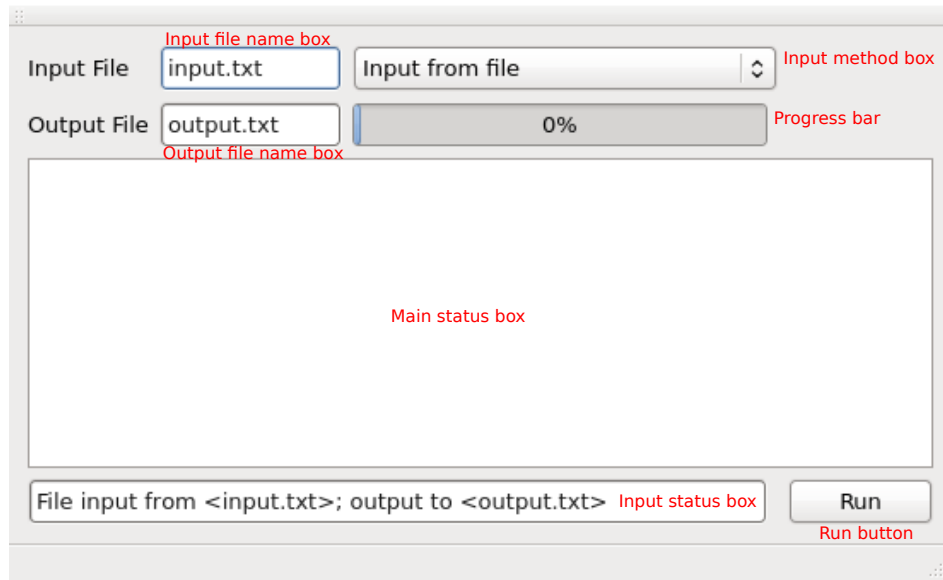


FIGURE 3. Default screen for MONOPHYLER upon startup with components labeled in red text.

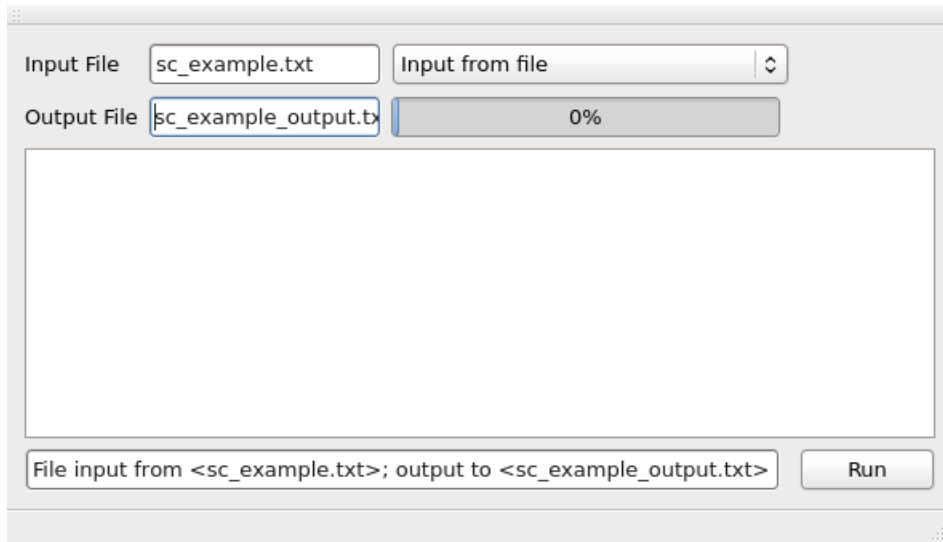


FIGURE 4. Changed file names to match example in Section 5.7.

example appears in Figure 7 with its components labeled. The manual input option can only be accessed from the GUI and not directly from the command line.

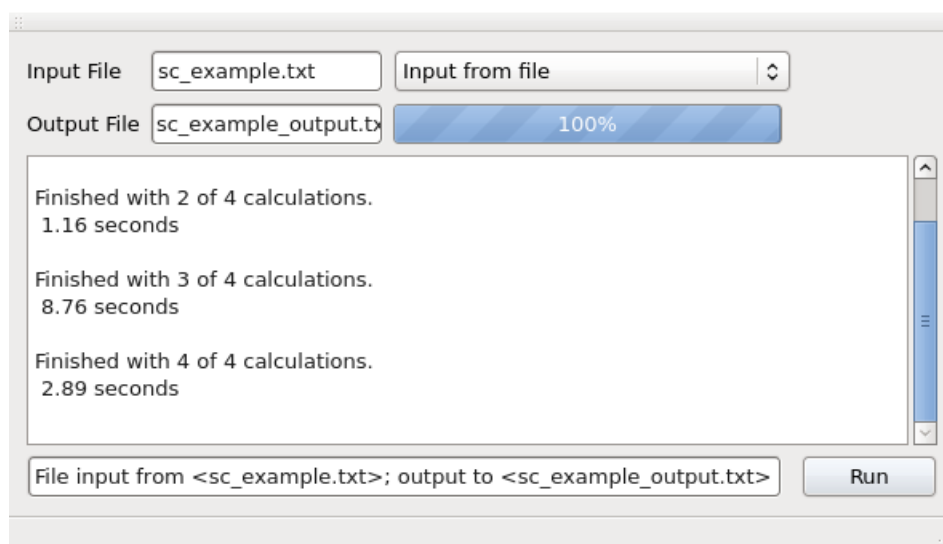


FIGURE 5. Screen of MONOPHYLER after the calculations from example in Section 5.7 are completed.

```

Line;Calculation;Number_of_Species;Tree;Initial_S_Lineages;Initial_C_Lineages;Branch_Lengths;Monophyly_Event;Probability
5;sc;2;(AB);{1,2};{3,1};{1,5001,2};S;0.00183857
6;sc;3;((AB)C);{1,2,3};{4,2,1};{1,1.5,2,5001,3};SC;1.45995e-07
7;sc;5;(((AB)(CD))E);{0,1,3,0,2};{2,0,1,3,4};{0.2,1,0.3,2,0.5,1.2,0.6,5001,5};C;7.64864e-09
8;sc;5;((AB)(C(DE)));{1,2,0,3,4};{0,0,1,2,1};{0.2,1,0.3,5001,0.5,2,0.6,1.2,5};S;7.57277e-08

```

FIGURE 6. Output file for example input file from Figure 1.

6.1. **Calculation menu.** The calculation menu lists the four calculation options seen in Table 1. You must choose one of these options. Choosing anything other than “subset-complement” will pre-populate some of the other boxes.

6.2. **Number of species box.** This box allows the user to set the number of species in the species tree. The user may choose any integer up to 100. All non-subset-complement calculations will set this box automatically, and the user cannot change the value.

6.3. **Tree input box.** This box is where the user types in the tree topology. The user must use the same format as in Section 5.2. However, the user may include or omit quotes here. This box is pre-populated in every non-subset-complement calculation and cannot be edited in those cases.

6.4. **S input box.** This box is where the user types in the sample sizes for the Class-*S* lineages and requires the same format as Section 5.3. This box is pre-populated with

Calculation menu

Monophyly of:

No. of species

Tree input box Tree

S input box S Sample Sizes

Branch length input box Branch Lengths

Populate with example

Number of species box

Example button

This box is only for the single-subset calculation.

C input box C Sample Sizes

Monophyly event: Monophyly event menu

Please input a tree in tree box above.

Tree status box

Please input S sample sizes in S sample size box above.

S status box

Please input branch lengths in branch lengths box above.

Branch length status box

Please input C sample sizes in C sample size box above.

C status box

Clear button Send button

Clear Send to main window

Main status box

FIGURE 7. Manual input window with components labeled in red.

the appropriate number of commas when any calculation other than subset-complement is selected.

6.5. Branch length box. This box is where the user types in branch lengths. This box requires the same format as Section 5.5. This box is pre-populated with the appropriate number of commas and the root branch length when any calculation other than subset-complement is selected.

6.6. **C input box.** This box is where the user types in the sample sizes for the Class- C lineages and requires the same format as Section 5.4. This box is only used in the subset-complement calculation and cannot be edited in any other situation.

6.7. **Monophyly event menu.** This menu allows the user to choose which monophyly event (S , C , or SC) is to be calculated. This menu is only used in the subset-complement calculation and is ignored by the program otherwise.

6.8. **Tree, S , branch length, and C status boxes.** These boxes indicate the validation status of their corresponding input boxes. If one of the inputs is incorrect, the status box will indicate the problem. If the input is correct, the status box will state “[input] verified”.

6.9. **Main status box.** This box will provide the status of the manual input calculation upon pressing the `Send to main window` button or the `Clear` button.

6.10. **Send button.** This button compiles all the information in the manual input window into a format that MONOPHYLER can interpret to perform the desired calculation and then sends this information to the main window of MONOPHYLER. Pressing `Run` in the main window will perform the calculation.

6.11. **Clear button.** This button resets the manual input window and erases any calculation previously sent to the main window. After pressing this button, pressing `Run` in the main window while still in `Manual input` mode will not perform any calculation.

6.12. **Example button.** This button populates the manual input window with an example calculation that corresponds to the choice of calculation in the calculation menu. If no calculation is chosen (i.e. the calculation menu states `Monophyly of:`) then the subset-complement example is used.

7. PERFORMING A MANUAL INPUT CALCULATION

This section will walk through an example manual input calculation. Running MONOPHYLER and then choosing `Manual input (opens in new window)` from the input method box in the main window will open the manual input window seen in Figure 7. Pressing `Populate with example` populates the manual input window with an example subset-complement calculation, with the result shown in Figure 8.

The calculation needs to be sent to the main window. Pressing `Send to main window` does this, with the result for the manual input window shown in Figure 9 and the result for the main window shown in Figure 10.

Subset-complement No. of species 3

((AB)C) Tree

3,0,4 S Sample Sizes Populate with example

0.3,0.1,1,5001,1.3 Branch Lengths

This section is only used by the single-subset calculation.

0,5,0 C Sample Sizes

S

Tree verified.

S sample sizes verified.

Branch lengths verified.

C sample sizes verified.

Clear Send to main window

FIGURE 8. Manual input window populated with subset-complement calculation example.

Manual input calculation results print to an output file in exactly the same way as file input calculations, except that manual input calculations do not overwrite the existing output file but rather append their results to any existing results in the output file. All settings are printed to the output file. manual input calculations will have line number 0 in the output file. An output file needs to be specified in the main window for the manual input calculation to occur. Pressing **Run** in the main window once the manual input has been sent to the main window will perform the calculation. The progress bar will not

Subset-complement No. of species 3

((AB)C) Tree

3,0,4 S Sample Sizes Populate with example

0.3,0.1,1,5001,1.3 Branch Lengths

This section is only used by the single-subset calculation.

0,5,0 C Sample Sizes

S

Tree verified.

S sample sizes verified.

Branch lengths verified.

C sample sizes verified.

Input sent to main window. Press <Run> in main window to perform the calculation.

Clear Send to main window

FIGURE 9. Manual input window populated with subset-complement calculation example, with input sent to main window

respond to a manual calculation. Figure 11 displays the result, which is also written to the output file (in this case `output.txt`).

If the user changes the output file name from file A to file B and then runs a manual input calculation, the result will be appended to file B. If the user then changes the file name back to file A and then runs a manual calculation, the result will be appended to file A.

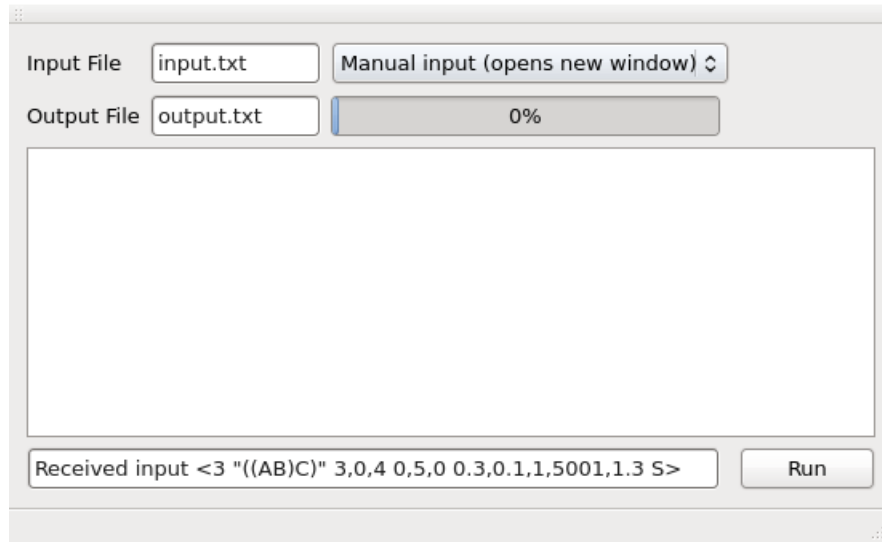


FIGURE 10. Main window with manual input example calculation sent from manual input window.

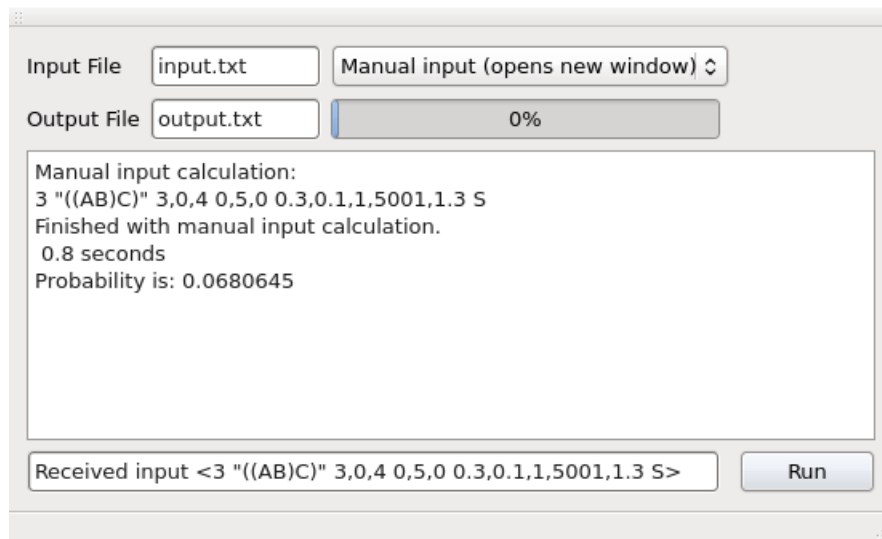


FIGURE 11. Main window after performing manual input calculation.

8. JOINT MONOPHYLY

The python script `joint_monophyler.py` computes the probability of joint monophyly for any set of species given a species tree, using the Continuous-Time Markov Chain

(CTMC) framework presented in Mehta, Steel, and Rosenberg (in prep). Running the script from the command line uses two options: ‘-m’ for manual input, and ‘-f’ for file input.

For manual input, additional arguments need to be presented in the command line. They are, in order: a newick-formatted string for the species tree, a comma-separated list of species names (it is sufficient/recommended to just use single characters), and a comma-separated list of sample sizes corresponding to each species. An example input line is

```
python joint_monophyler.py -m (A:0.0019,(B:0.0004,C:0.00035):0.0016); A,B,C
2,6,3
```

For file input, the command line should include, in order: the file name for the species trees, the file name for a list of sample names, the file name for a list of sample sizes, and the file name for the output. An example input line is:

```
python joint_monophyler.py -f trees.newick sample_names.txt sample_sizes.txt
output_name
```

The files are space-delimited. An example tree file is:

```
(A:0.0019,(B:0.0004,C:0.00035):0.0016); (A:0.2,(C:0.4,B:0.3):0.1); (A:0.0016,(D:0.0003,F:0.0001):0.006);
```

An example sample name file is:

```
A,B,C A,C,B A,D,F
```

An example sample size file is:

```
2,3,7 6,2,3 8,2,5
```

9. ERRORS TO IGNORE

Running MONOPHYLER through a command line may cause certain error messages to be displayed even if the program is working correctly. The user may ignore the following error messages if they occur as they do not have anything to do with the actual running of the software:

- Qt: Session management error: None of the authentication protocols specified are supported
- QObject::connectSlotsByName: No matching signal for on_manual_input_window_closed()
- QObject::connectSlotsByName: No matching signal for on_manual_input_relayed()
- QObject::connectSlotsByName: No matching signal for on_manual_input_window_cleared()
- Xlib: extension "GLX" missing on display ":0.0".
- Could not initialize OpenGL for RasterGLSurface, reverting to RasterSurface.

The last two errors may be important if the user’s computer is unable to display the GUI.

10. TROUBLESHOOTING

This section details some possible errors that may not be obvious from the provided error message (or may result in the program crashing and giving the user no information as to what happened).

- (1) Extraneous nonnumeric characters to start a post-header line will yield an error message unless the character in question is a single or double quote. A single quote (') will cause the program to crash (and an error message indicating an unterminated string may occur). A double quote (") will do the same if there is no other double quote in the same line in the file (which means the tree topology is not formatted correctly). If there is a double quote somewhere in the rest of the line then the line will also yield an error message.
- (2) Any single quotes anywhere in the input file will cause a runtime error (possibly with a message indicating an unterminated string).
- (3) Stray double quotes may also result in a runtime error.
- (4) Extraneous nonnumeric characters between the number of species and the tree topology in an input file will be interpreted as the tree topology itself and will fail one of the tree topology format checks.

11. UPDATES

Version 1.1 - Added libraries to Linux package that resolve the following error on certain Linux machines:

```
This application failed to start because it could not find or load the Qt platform
plugin "xcb".
```

Version 2.0 - Added `joint_monophyler.py`.

12. REFERENCES

Mehta, Rohan S., David Bryant, and Noah A. Rosenberg. "The probability of monophyly of a sample of gene lineages on a species tree." *Proc Natl Acad Sci USA* 113 (2016): 8002-8009.

Mehta, Rohan S. and Noah A. Rosenberg. "The probability of reciprocal monophyly for three and four species." *Theor Popul Biol* 129 (2019): 123-147.

Mehta, Rohan S., Mike Steel, and Noah A. Rosenberg. "The probability of joint monophyly of all species in an arbitrary species tree." In prep.

Rosenberg, Noah A. "The probability of topological concordance of gene trees and species trees." *Theor Popul Biol* 61 (2002): 225-247.

Rosenberg, Noah A. "The shapes of neutral gene genealogies in two species: probabilities of monophyly, paraphyly, and polyphyly in a coalescent model." *Evolution* 57 (2003): 1465-1477.

```

Input file for Monophyler version 1.0 7/21/16
4
2
END HEADER
2 "(AB)" 1,2 1,5001,2
2 "(AB)" 2,2 0,5001,2
2 "(AB)" 1,3 1,5001,0.1
2 "(AB)" 5,5 1,5001,2

```

FIGURE 12. Example input file 2rm_example.txt.

```

Input file for Monophyler version 1.0 7/21/16
4
3
END HEADER
3 "((AB)C)" 1,2,3 1,1.5,2,5001,3
3 "((AB)C)" 4,2,1 2,1.5,2,5001,1
3 "((AB)C)" 1,1,1 1,1.5,1.5,5001,3
3 "((AB)C)" 5,5,5 0.2,0.2,0.2,5001,0.867

```

FIGURE 13. Example input file 3rm_example.txt.

```

Input file for Monophyler version 1.0 7/21/16
4
4
END HEADER
4 "((AB)(CD))" 1,2,3,5 0.2,0.1,1,5001,1,0.2,0.5
4 "((AB)(CD))" 2,2,4,4 0.2,0.2,0.2,5001,1,0.2,0.2
4 "((AB)(CD))" 1,2,3,1 0.2,0.1,1,5001,1,0.2,1
4 "((AB)(CD))" 5,5,5,5 1,0.1,1,5001,1,0.2,0.5

```

FIGURE 14. Example input file 4rmb_example.txt.

```

Input file for Monophyler version 1.0 7/21/16
4
5
END HEADER
4 "(((AB)C)D)" 1,2,3,5 0.2,0.1,1,1,0.2,5001,8
4 "(((AB)C)D)" 1,2,3,1 0.2,0.1,1,1,0.2,5001,0.5
4 "(((AB)C)D)" 2,2,2,2 2,2,2,2,5001,2
4 "(((AB)C)D)" 5,5,5,5 0.2,0.001,1,1,0.2,5001,0.5

```

FIGURE 15. Example input file 4rmc_example.txt.