

MEGA-CC (Compute Core) and MEGA-Proto

Quick Start Tutorial

Installation

- Create a folder for MEGA (*i.e.* `~\Documents\MEGA-CC`).
- Copy *M6CC.exe* and *M6Proto.exe* to the newly created directory.
- For the demo at the end of this document, also copy the *Examples* folder that came with the M6CC.zip file. This folder contains the input data file that will be used for the demo.

MEGA-CC Input Files

- MEGA Analysis Options file
 - Specifies the calculation and desired settings.
 - Created using MEGA-Proto.
 - Has a *.mao* file extension.
- Data file (one of the following)
 - Multiple sequence alignment in MEGA or Fasta format.
 - Distance matrix in MEGA format.
 - Unaligned sequences in Fasta format (for alignment only).
- Tree file (required for some analyses)
 - Newick file format.

MEGA-CC Output Files

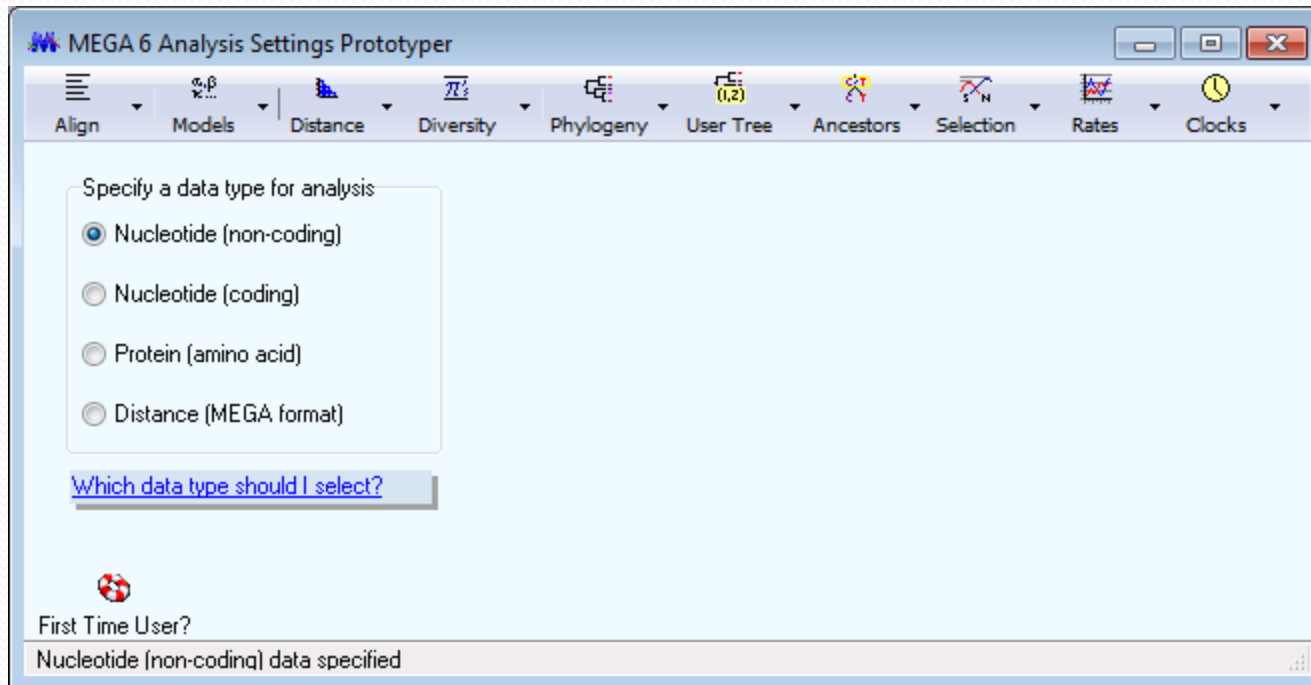
- In general, two output files are produced
 1. Calculation-specific results file (Newick file, distance matrix,...).
 2. A summary file with additional info (likelihood, SBL,...).
- Some analyses produce additional output (bootstrap consensus tree).
- Output directory/filename
 - Default is the same location as the input data file.
 - Specify an output directory and/or file name using -o option.
 - If no output filename is specified, MEGA-CC will assign a unique name.
- Errors/warnings
 - If MEGA-CC produces any errors or warnings, they will be logged in the the summary file.

Running MEGA-CC

- Easiest to run using command-line or batch scripts:
 - `M6CC.exe -a options.mao -d alignment.meg -o outFile`
- Can also be run using custom scripts (Perl, Python, ...):
 - `exec('M6CC.exe -a options.mao -d alignment.meg -o outFile');`
- Integrated *File Iterator* system can process multiple files without the need for using scripts (see Demo2 below)
- In addition, other applications can launch MEGA-CC:
 - `status = CreateProcess("M6CC.exe...");`
- To see a list of available command options, call `M6CC.exe` from a command-line prompt with the `-h` flag.

MEGA-Proto (analysis prototyper)

- Has the same look and feel as the GUI edition of MEGA.
- Produces MEGA Analysis Options files.
- Has no computational capabilities.

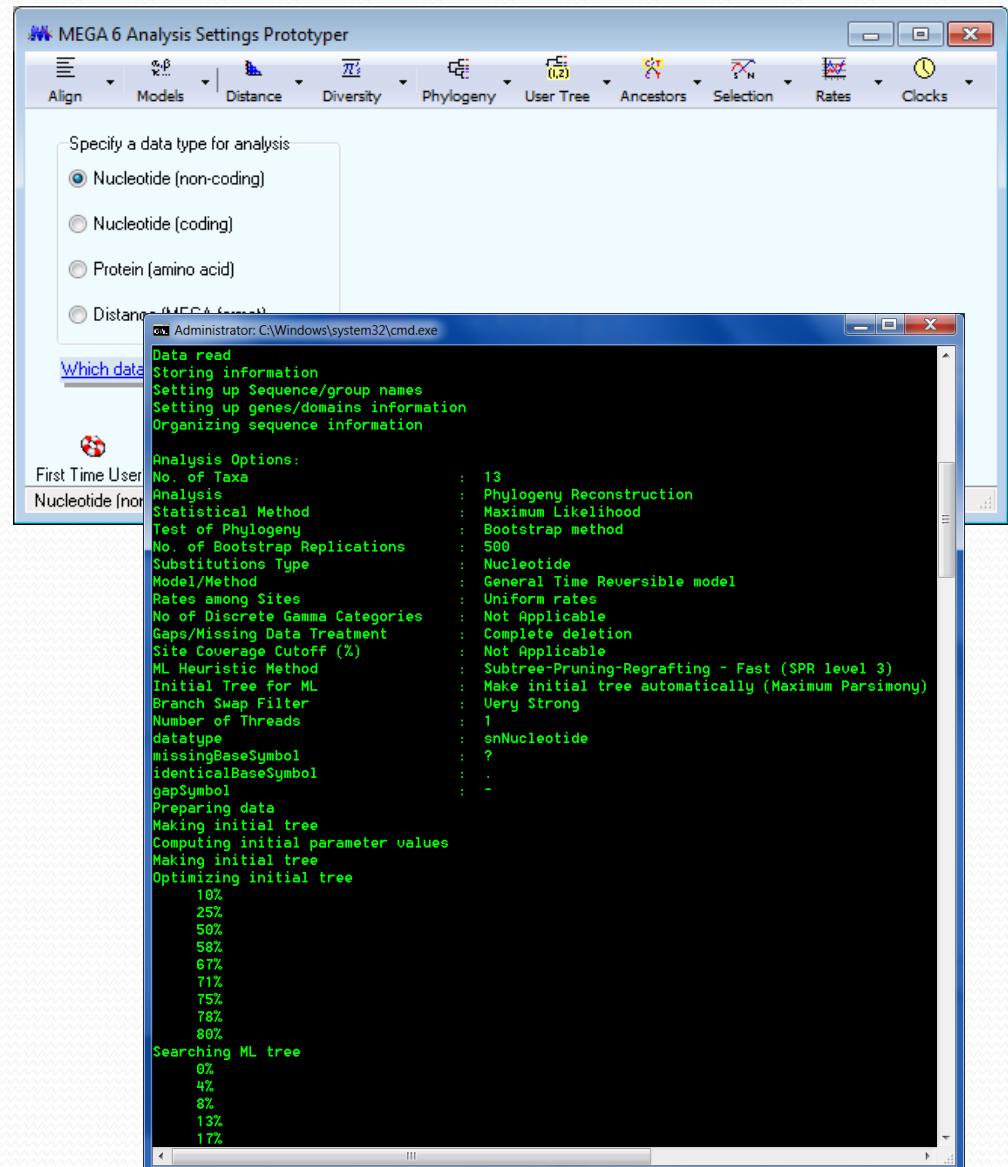


Using MEGA-Proto

1. Select input data type.
 - Nucleotide (non-coding)
 - Nucleotide (coding)
 - Protein (amino-acid)
 - Distance matrix (MEGA format)
2. Select analysis from menu.
3. Adjust analysis settings.
4. Save the MEGA Analysis Options file.

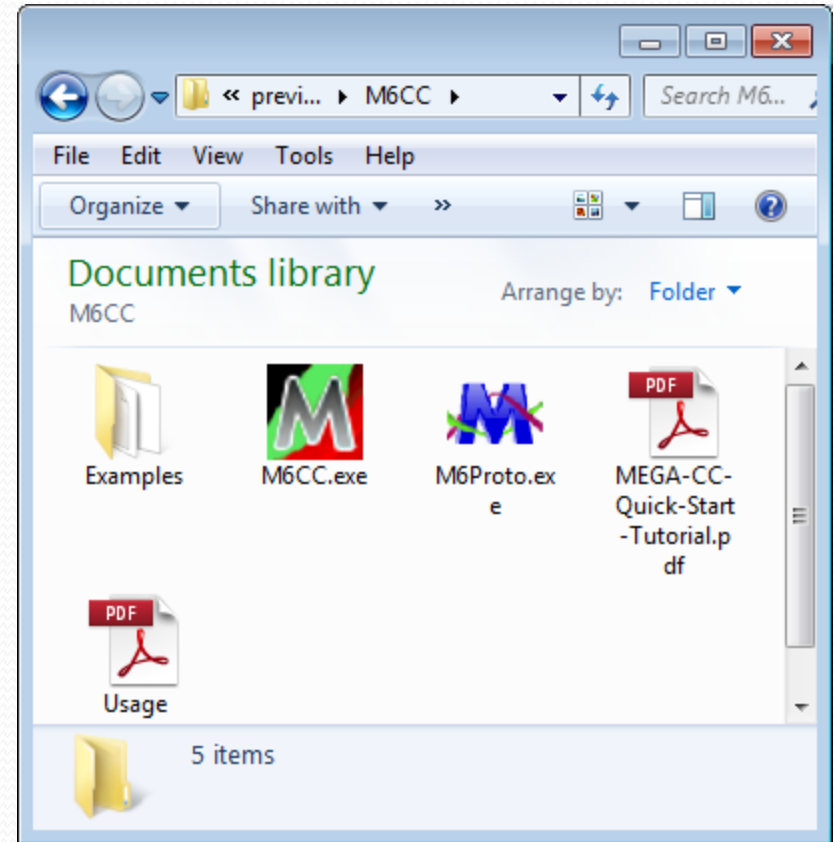
Demo1

- The following example demonstrates how to perform Maximum Likelihood phylogeny reconstruction using MEGA-Proto and MEGA-CC



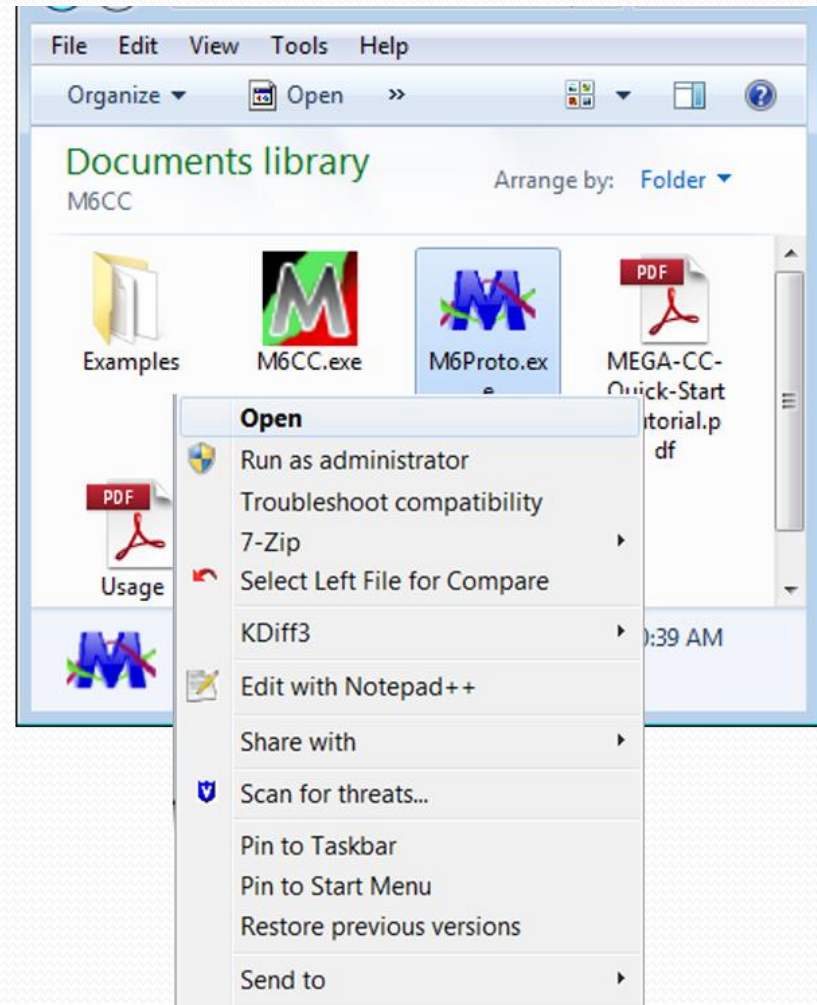
Demo1 Setup

- For this demo, make sure that you have the M6CC.exe, M6Proto.exe, and example data files located in your working directory as shown. The example data files folder is included in the MEGA-CC.zip download.



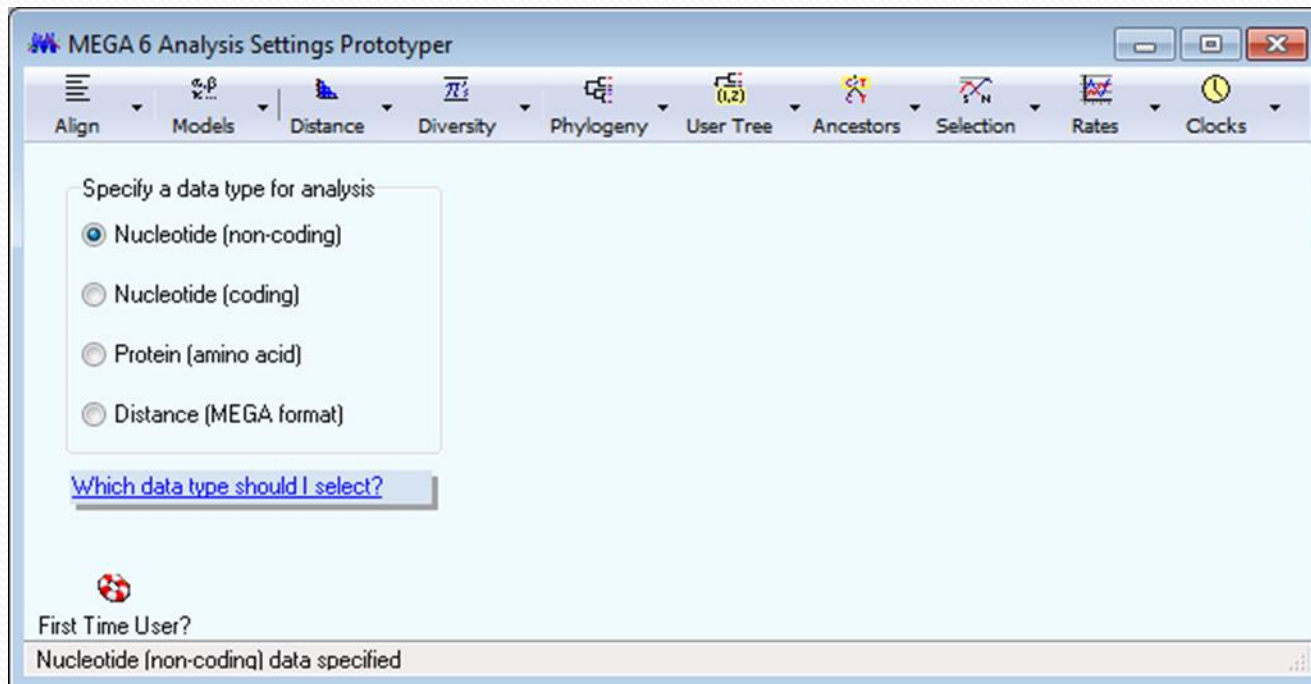
Step 1

- Open MEGA-Proto by right-clicking the M6Proto.exe file and selecting *Open*.



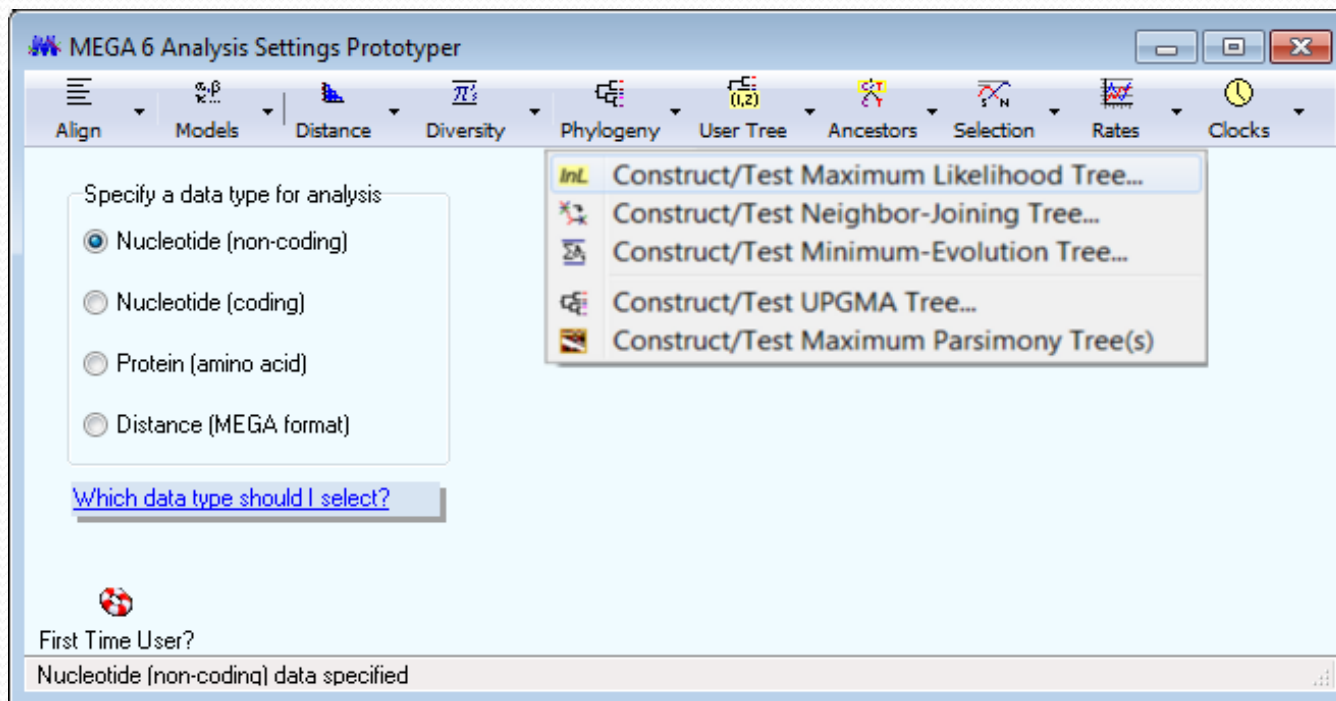
Step 2

- Select the data type of the input data file to be analyzed. For this demo, we will accept the default setting - Nucleotide (non-coding).



Step 3

- Select *Construct/Test Maximum Likelihood Tree...* from the *Phylogeny* menu.



Step 4

- Adjust the analysis preferences to match those shown.
- Click the *Save Settings...* button and save the analysis options file as *mlDemo.mao* in the current working directory.

M6: Analysis Preferences

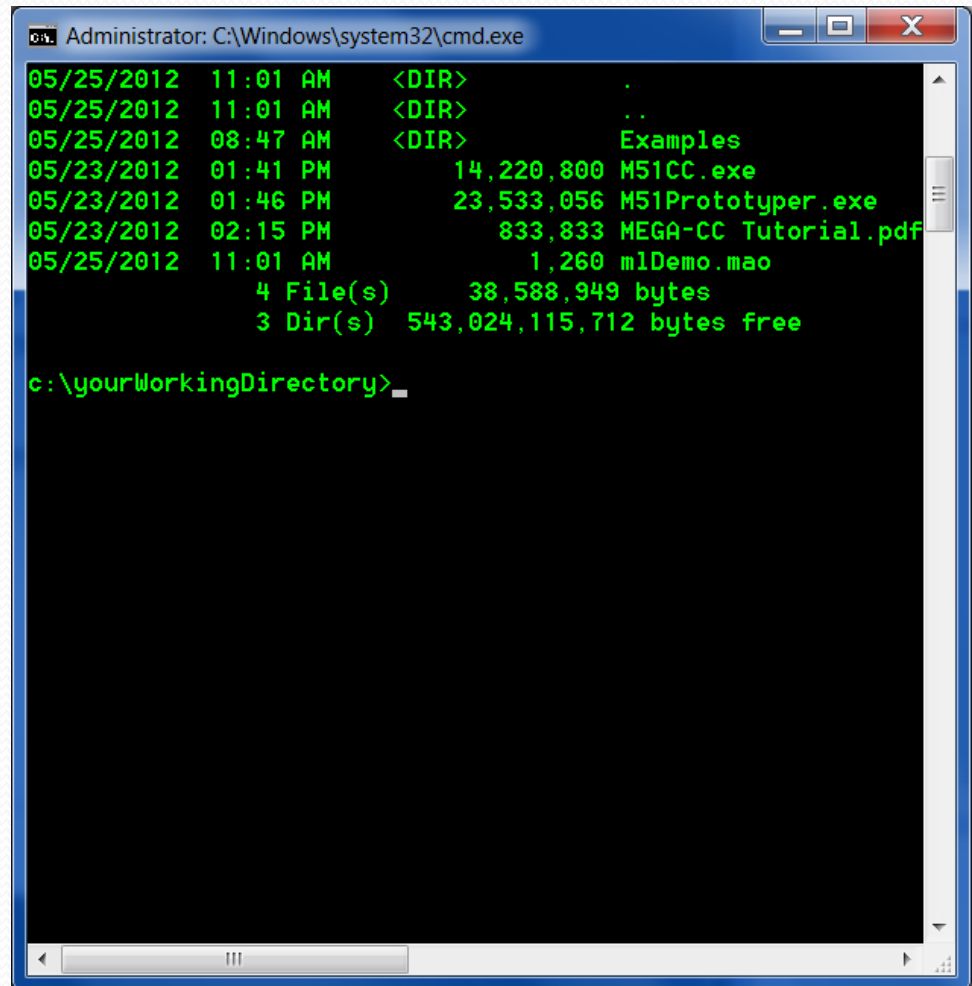
Options Summary | Gaps/Identical/Missing Data Treatment

Option	Selection
Analysis	Phylogeny Reconstruction
Statistical Method	Maximum Likelihood
Phylogeny Test	
Test of Phylogeny	Bootstrap method
No. of Bootstrap Replications	500
Substitution Model	
Substitutions Type	Nucleotide
Model/Method	General Time Reversible model
Rates and Patterns	
Rates among Sites	Uniform rates
No. of Discrete Gamma Categories	Not Applicable
Data Subset to Use	
Gaps/Missing Data Treatment	Complete deletion
Site Coverage Cutoff (%)	Not Applicable
Tree Inference Options	
ML Heuristic Method	Subtree-Pruning-Regrafting - Fast (SPR level 3)
Initial Tree for ML	Make initial tree automatically (Maximum Parsimony)
Branch Swap Filter	Very Strong
System Resource Usage	
Number of Threads	1

Save Settings... X Cancel

Step 5

- Open a command prompt.
- Navigate to the same directory as the M6CC.exe file using the *cd* command



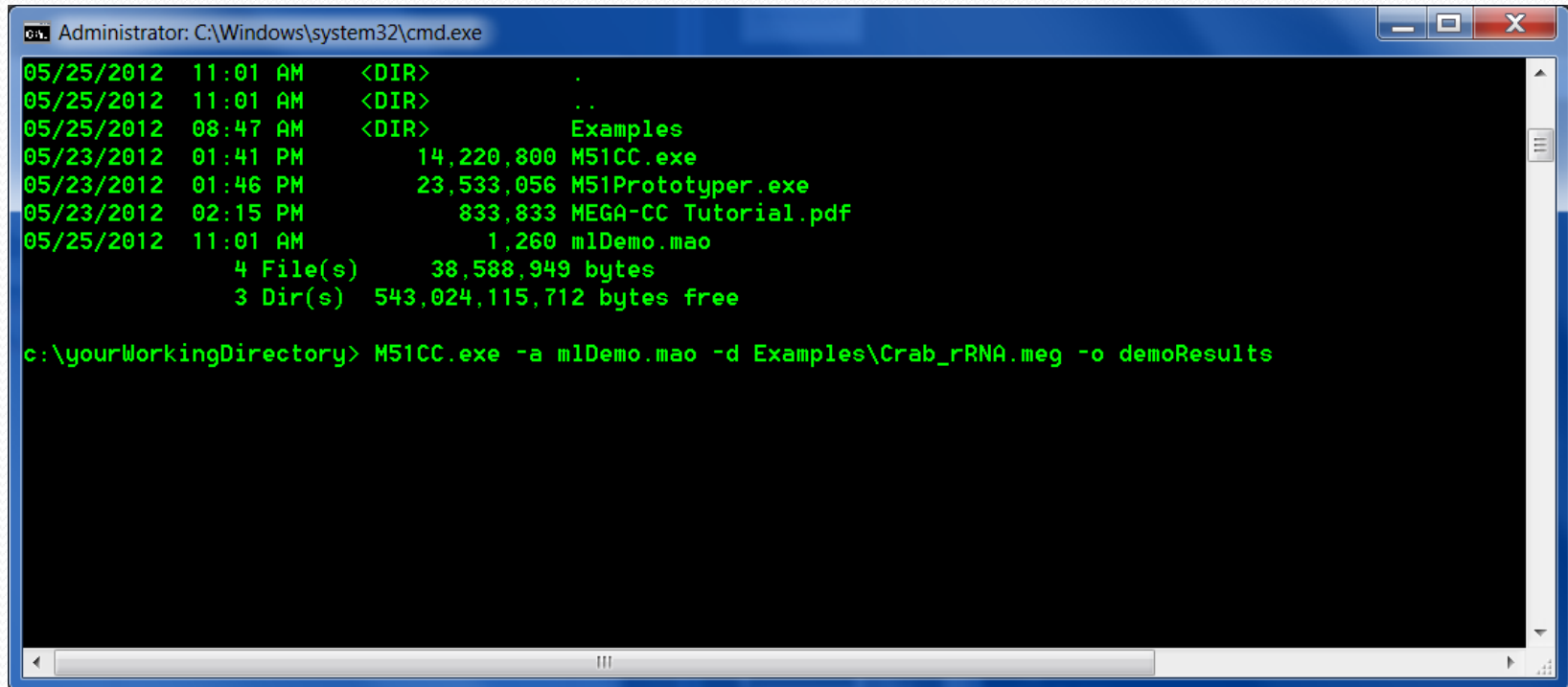
A screenshot of a Windows command prompt window titled "Administrator: C:\Windows\system32\cmd.exe". The window displays the output of a `DIR` command, showing a list of files and directories. The output is as follows:

```
05/25/2012 11:01 AM <DIR> .
05/25/2012 11:01 AM <DIR> ..
05/25/2012 08:47 AM <DIR> Examples
05/23/2012 01:41 PM      14,220,800 M51CC.exe
05/23/2012 01:46 PM      23,533,056 M51Prototyper.exe
05/23/2012 02:15 PM       833,833 MEGA-CC Tutorial.pdf
05/25/2012 11:01 AM        1,260 m1Demo.mao
                   4 File(s)      38,588,949 bytes
                   3 Dir(s)  543,024,115,712 bytes free

c:\yourWorkingDirectory>
```

Step 6

- Execute the analysis by calling the M6CC.exe file
 - M6CC.exe -a mlDemo.mao -d Examples\Crab_rRNA.meg -o demoResults



The screenshot shows a Windows command prompt window titled "Administrator: C:\Windows\system32\cmd.exe". The window displays the output of a `DIR` command, listing files and directories in the current path. The files listed are `M51CC.exe`, `M51Prototyper.exe`, `MEGA-CC Tutorial.pdf`, and `mlDemo.mao`. The command prompt then shows the execution of the command `c:\yourWorkingDirectory> M51CC.exe -a mlDemo.mao -d Examples\Crab_rRNA.meg -o demoResults`.

```
Administrator: C:\Windows\system32\cmd.exe
05/25/2012  11:01 AM    <DIR>          .
05/25/2012  11:01 AM    <DIR>          ..
05/25/2012  08:47 AM    <DIR>          Examples
05/23/2012  01:41 PM             14,220,800 M51CC.exe
05/23/2012  01:46 PM             23,533,056 M51Prototyper.exe
05/23/2012  02:15 PM              833,833 MEGA-CC Tutorial.pdf
05/25/2012  11:01 AM              1,260 mlDemo.mao
               4 File(s)      38,588,949 bytes
               3 Dir(s)    543,024,115,712 bytes free

c:\yourWorkingDirectory> M51CC.exe -a mlDemo.mao -d Examples\Crab_rRNA.meg -o demoResults
```

Step 7

- The analysis will be launched and progress updates will be displayed in the command prompt window.

```
Administrator: C:\Windows\system32\cmd.exe

c:\yourWorkingDirectory> M51CC.exe -a mlDemo.mao -d Examples\Crab_rRNA.meg -o demoResults
MEGA 5.1 Molecular Evolutionary Genetics Analysis
Build#: 5120301
Data file           : Examples\Crab_rRNA.meg
Reading header
Reading data
Data read
Storing information
Setting up Sequence/group names
Setting up genes/domains information
Organizing sequence information

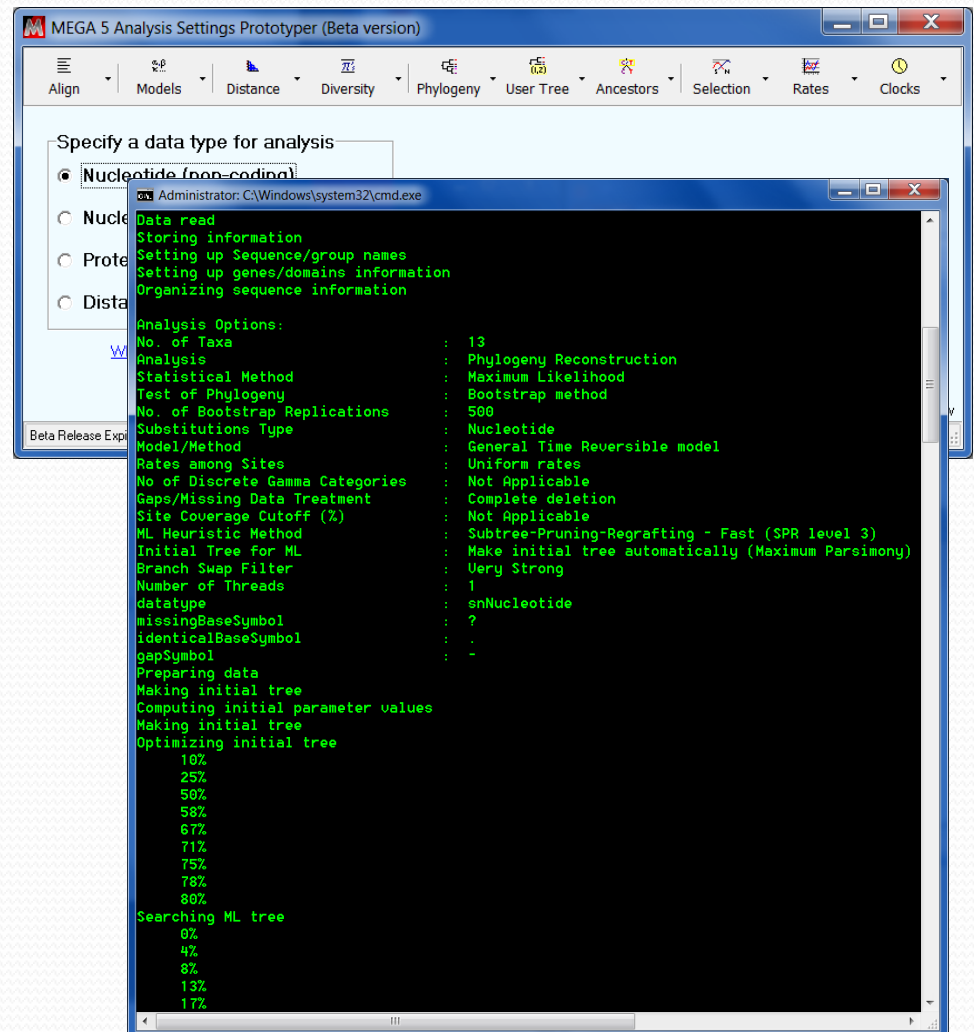
Analysis Options:
No. of Taxa           : 13
Analysis              : Phylogeny Reconstruction
Statistical Method    : Maximum Likelihood
Test of Phylogeny     : Bootstrap method
No. of Bootstrap Replications : 500
Substitutions Type    : Nucleotide
Model/Method          : General Time Reversible model
Rates among Sites     : Uniform rates
No of Discrete Gamma Categories : Not Applicable
Gaps/Missing Data Treatment : Complete deletion
Site Coverage Cutoff (%) : Not Applicable
ML Heuristic Method   : Subtree-Pruning-Regrafting - Fast (SPR level 3)
Initial Tree for ML   : Make initial tree automatically (Maximum Parsimony)
Branch Swap Filter    : Very Strong
Number of Threads     : 1
datatype              : snNucleotide
missingBaseSymbol     : ?
identicalBaseSymbol   : .
gapSymbol              : -
Preparing data
Making initial tree
Computing initial parameter values
Making initial tree
Optimizing initial tree
10%
25%
50%
58%
67%
71%
75%
```


Finally

- The analysis will produce 3 output files
 - demoResults.nwk
 - This Newick file gives the Maximum Likelihood tree found using the given settings.
 - demoResults_consensus.nwk
 - This Newick file gives the bootstrap consensus tree that MEGA produces from all bootstrap sample trees.
 - demoResults_summary.txt
 - This file gives analysis information such as the log likelihood value of the Maximum Likelihood tree, ts/tv ratio, etc...

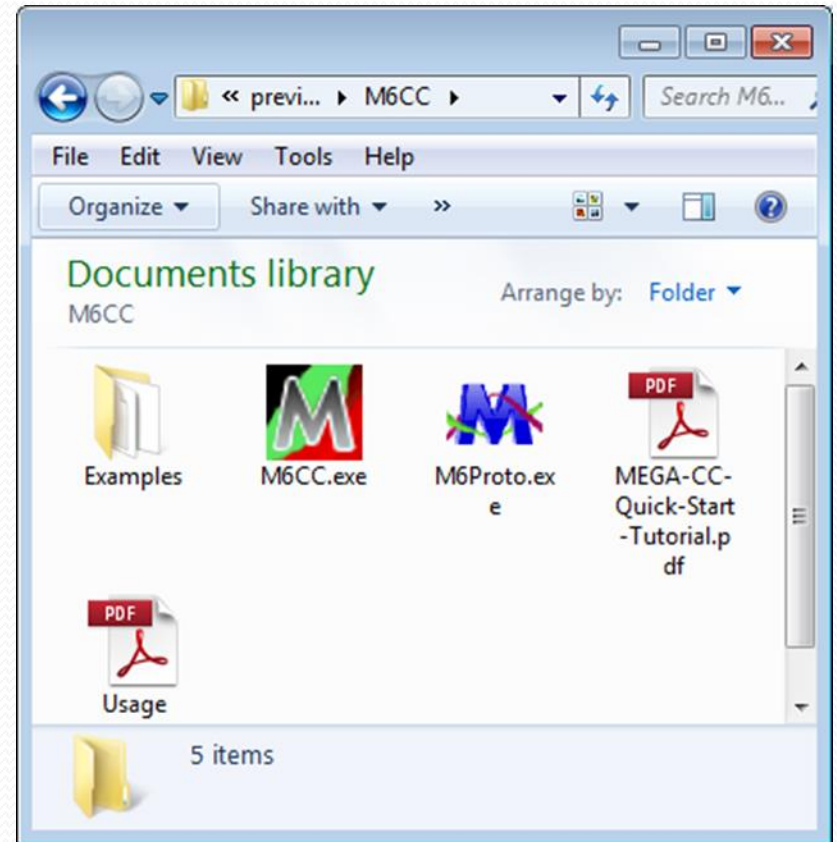
Demo2

- The following example demonstrates how to use the File Iterator system in MEGA-CC to process multiple input data files using a single analysis options file.



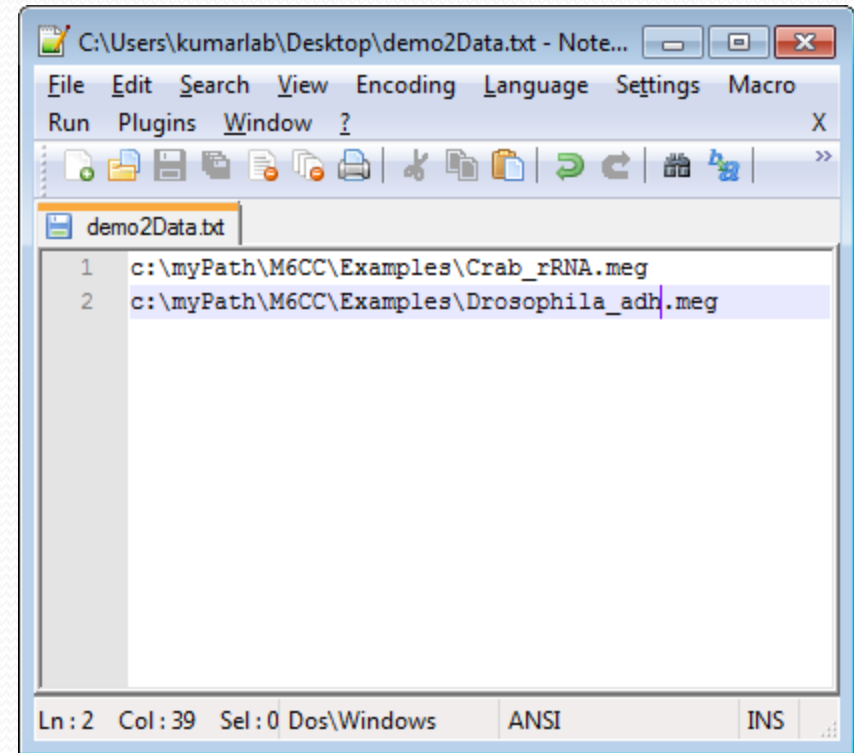
Demo2 Setup

- If you have not already done so, perform the Demo1 Setup as well as the first 5 steps of Demo1.



Step 1

- Create a text file named demo2Data.txt which we will use to specify multiple alignment files for ML phylogeny inference.
- In this file, add the full paths to the Crab_rRNA.meg and Drosophila_Adh.meg example files.

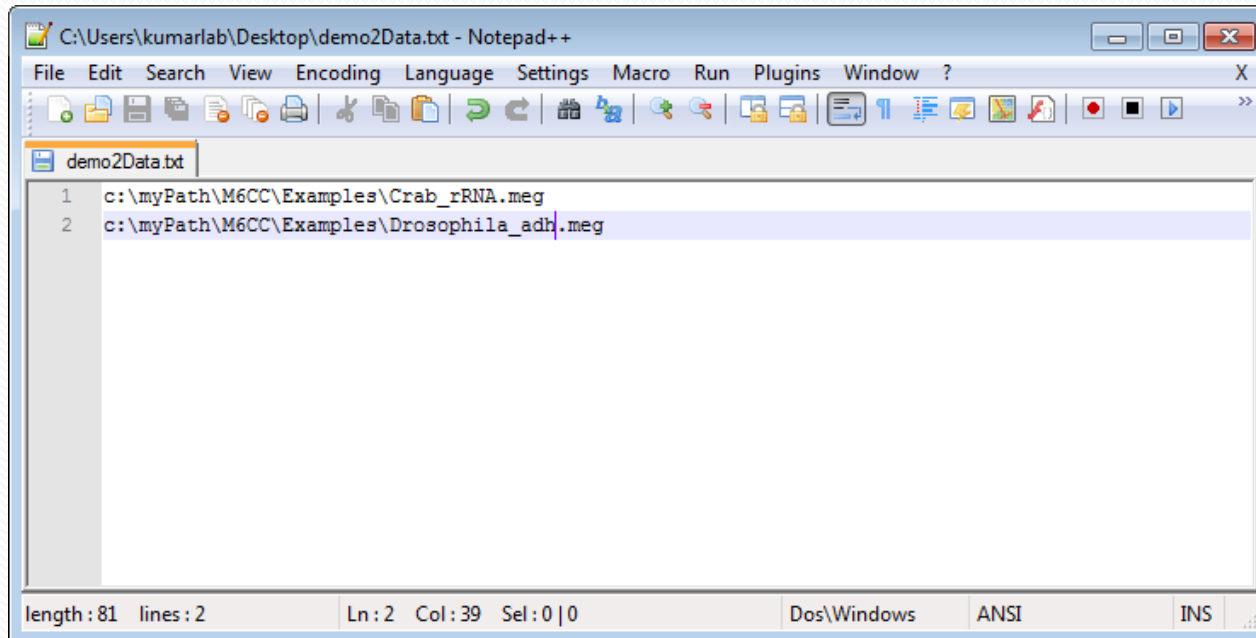


The screenshot shows a Notepad++ window titled "C:\Users\kumarlab\Desktop\demo2Data.txt - Note...". The window has a menu bar with "File", "Edit", "Search", "View", "Encoding", "Language", "Settings", and "Macro". Below the menu bar is a toolbar with various icons. The text area contains two lines of text, numbered 1 and 2 in the left margin. Line 1 is "c:\myPath\M6CC\Examples\Crab_rRNA.meg" and line 2 is "c:\myPath\M6CC\Examples\Drosophila_adh.meg". The status bar at the bottom shows "Ln : 2 Col : 39 Sel : 0 Dos\Windows ANSI INS".

```
1 c:\myPath\M6CC\Examples\Crab_rRNA.meg
2 c:\myPath\M6CC\Examples\Drosophila_adh.meg
```

Step 2

- From a command-line prompt, call MEGA-CC as follows:
 - M6CC.exe -a mlDemo.mao -d demo2Data.txt



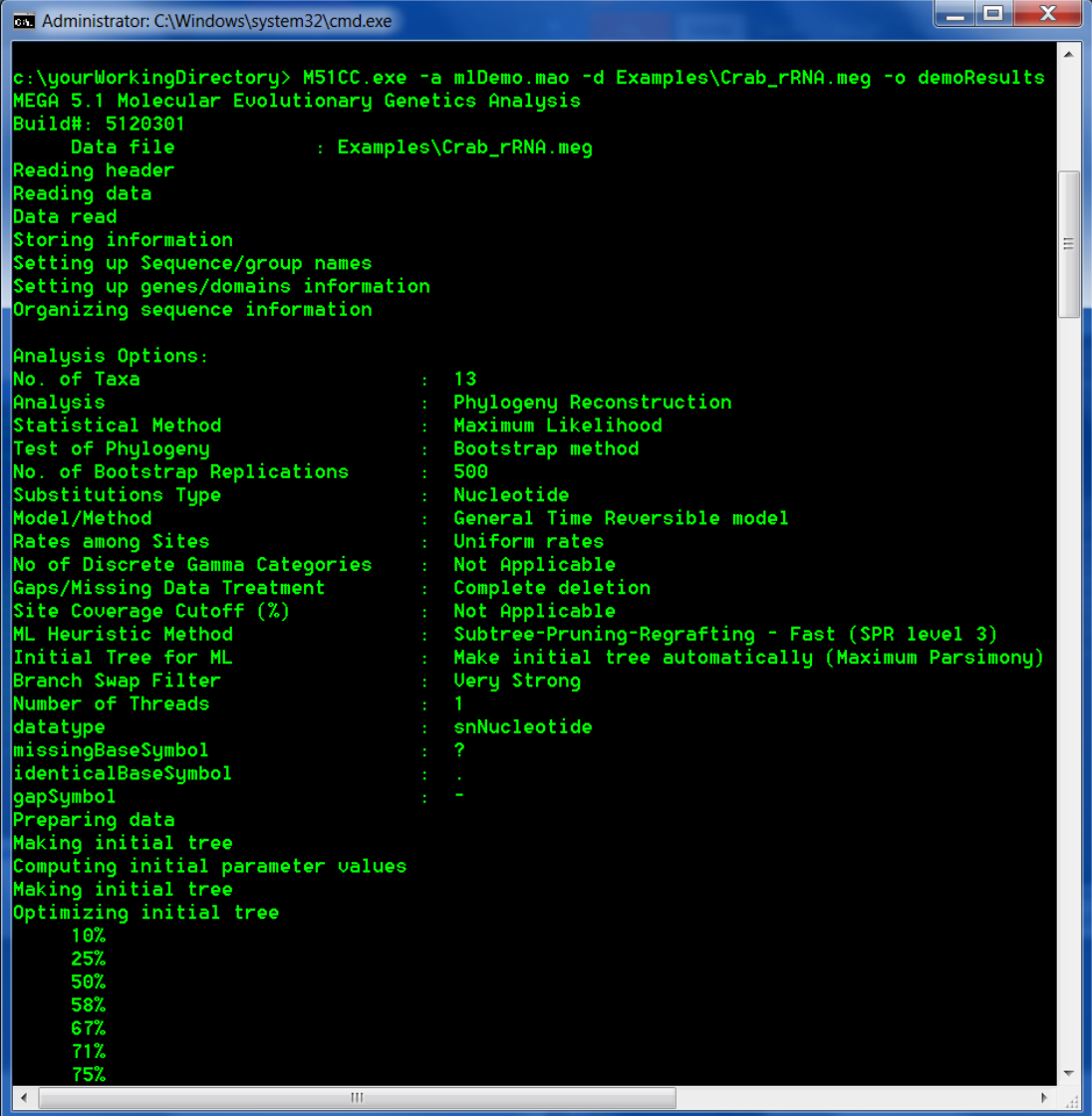
The screenshot shows a Notepad++ window titled "C:\Users\kumarlab\Desktop\demo2Data.txt - Notepad++". The window has a menu bar with "File", "Edit", "Search", "View", "Encoding", "Language", "Settings", "Macro", "Run", "Plugins", "Window", and "?". Below the menu bar is a toolbar with various icons. The main text area shows two lines of text:

```
1 c:\myPath\M6CC\Examples\Crab_rRNA.meg
2 c:\myPath\M6CC\Examples\Drosophila_adh.meg
```

The status bar at the bottom indicates "length : 81 lines : 2", "Ln : 2 Col : 39 Sel : 0 | 0", "Dos\Windows", "ANSI", and "INS".

Step 3

- The analyses will be launched sequentially and progress updates will be displayed in the command prompt window.



```
Administrator: C:\Windows\system32\cmd.exe

c:\yourWorkingDirectory> M51CC.exe -a mlDemo.mao -d Examples\Crab_rRNA.meg -o demoResults
MEGA 5.1 Molecular Evolutionary Genetics Analysis
Build#: 5120301
Data file           : Examples\Crab_rRNA.meg
Reading header
Reading data
Data read
Storing information
Setting up Sequence/group names
Setting up genes/domains information
Organizing sequence information

Analysis Options:
No. of Taxa           : 13
Analysis              : Phylogeny Reconstruction
Statistical Method    : Maximum Likelihood
Test of Phylogeny      : Bootstrap method
No. of Bootstrap Replications : 500
Substitutions Type     : Nucleotide
Model/Method          : General Time Reversible model
Rates among Sites      : Uniform rates
No of Discrete Gamma Categories : Not Applicable
Gaps/Missing Data Treatment : Complete deletion
Site Coverage Cutoff (%) : Not Applicable
ML Heuristic Method    : Subtree-Pruning-Regrafting - Fast (SPR level 3)
Initial Tree for ML    : Make initial tree automatically (Maximum Parsimony)
Branch Swap Filter     : Very Strong
Number of Threads      : 1
datatype              : snNucleotide
missingBaseSymbol      : ?
identicalBaseSymbol    : .
gapSymbol              : -
Preparing data
Making initial tree
Computing initial parameter values
Making initial tree
Optimizing initial tree
10%
25%
50%
58%
67%
71%
75%
```

Finally

- The analysis will produce output files for each input data file
- In this example, the same analysis options were used for each alignment file
- Enjoy!