MEGA-CC (Compute Core) and MEGA-Proto

Quick Start Tutorial

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

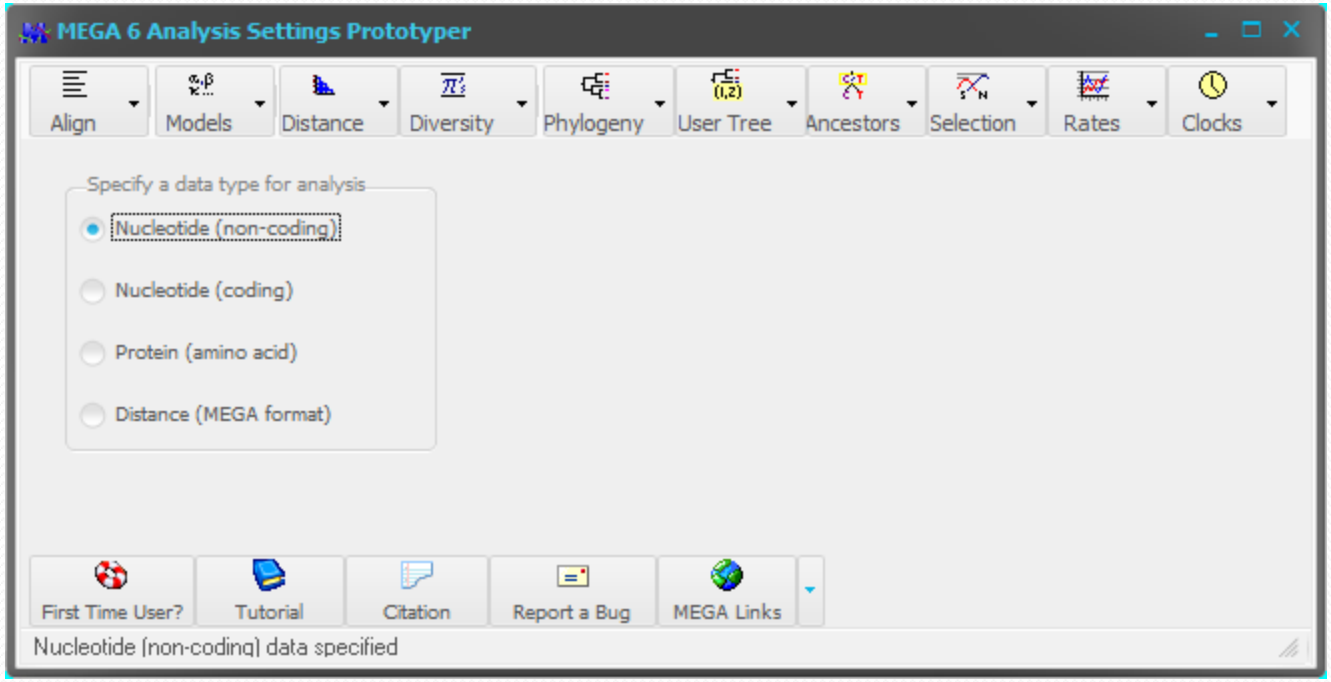
- In general, two output files are produced
  1. Calculation-specific results file (Newick file, distance matrix,…).
  2. A summary file with additional info (likelihood, SBL,…).
  - Some analyses produce additional output (bootstrap consensus tree).
- Output directory/filename
  - Default is the same 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:
  - megacc –a settings.mao –d alignment.meg –o outFile
- Can also be run using custom scripts (Perl, Python, …):
  - exec('megacc –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("path/to/megacc.exe…");
- To see a list of available command options, call megacc 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.

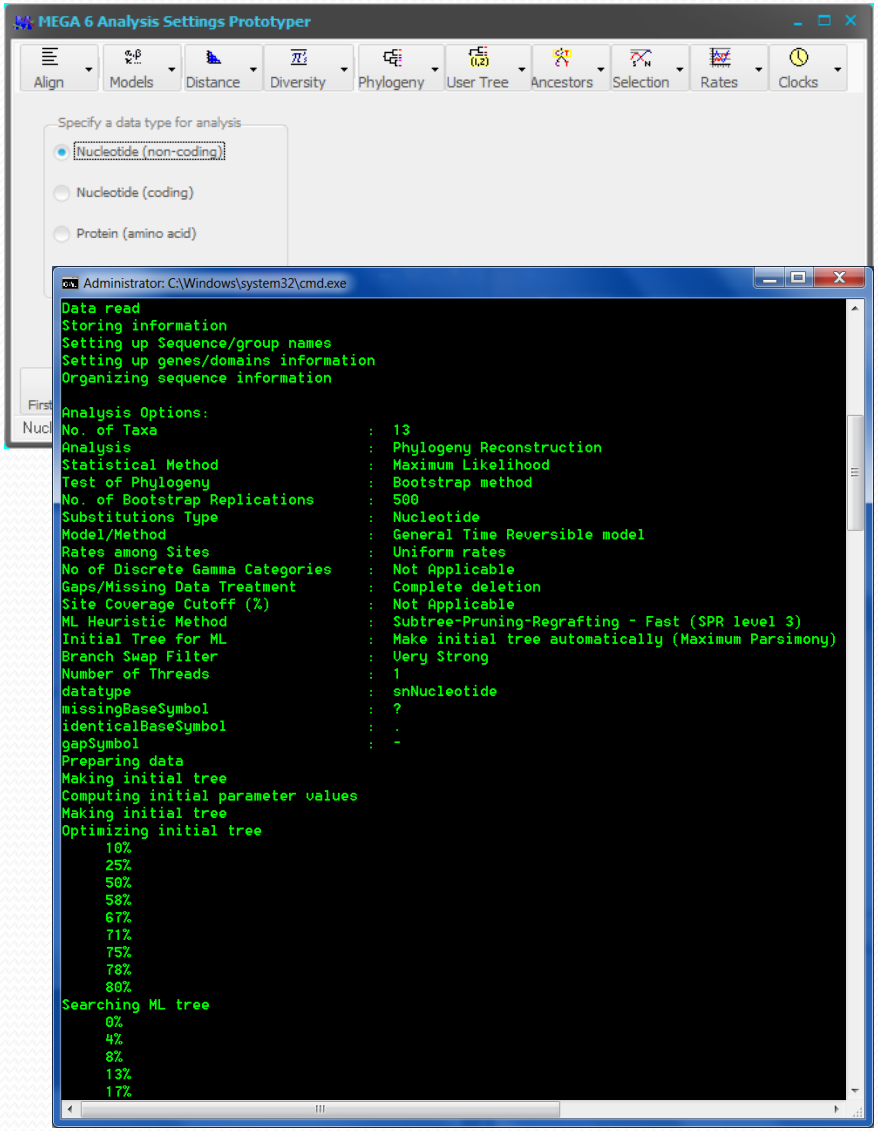


Using MEGA-Proto

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.

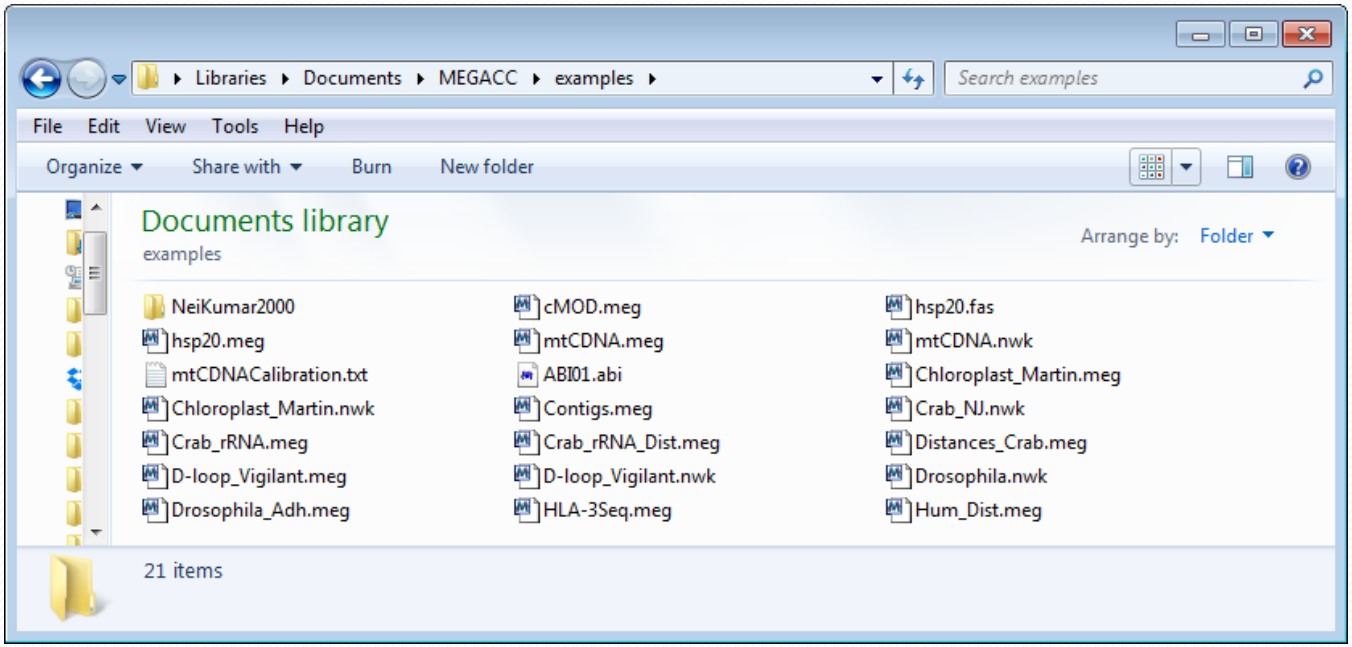
Demo1

- The following example demonstrates how to create a timetree using MEGA-Proto and MEGA-CC



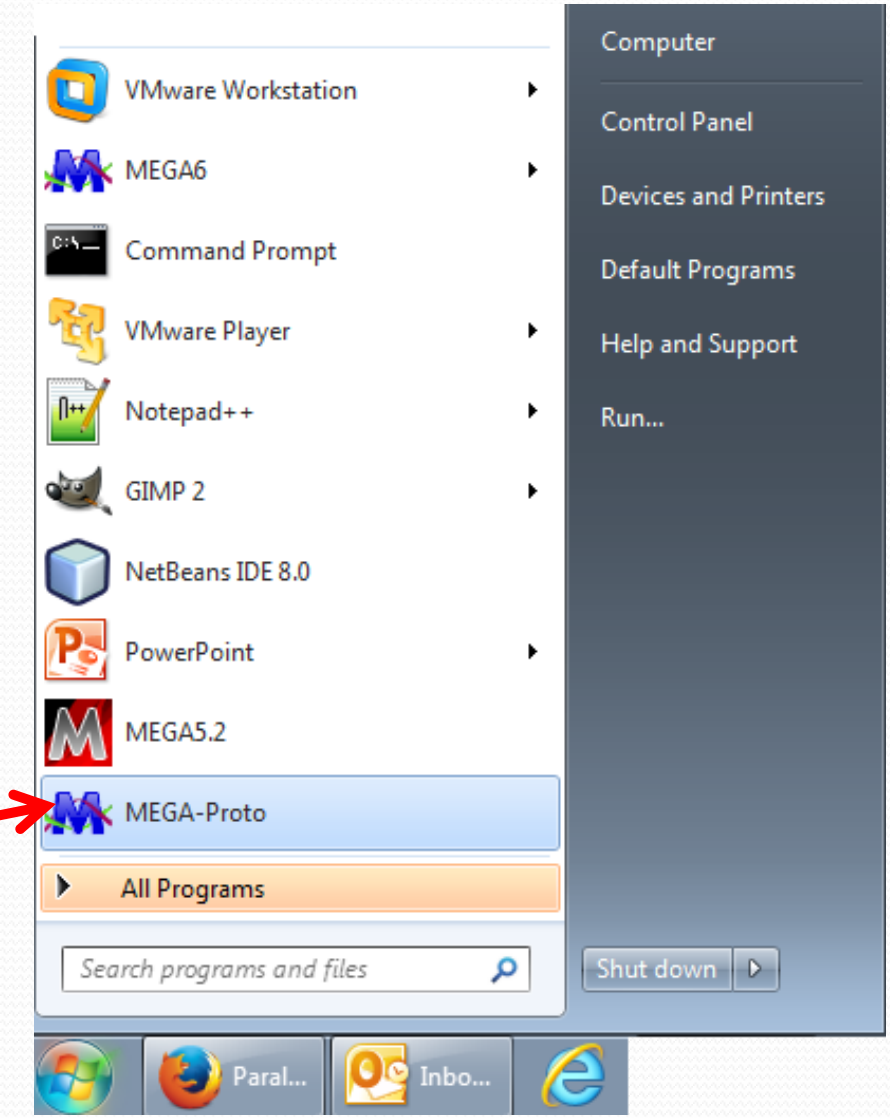
Demo1 Data Files

- For this demo, we will be using some of the example data files that were copied to your documents directory by the installer (*i.e.* Documents\MEGACC\examples).



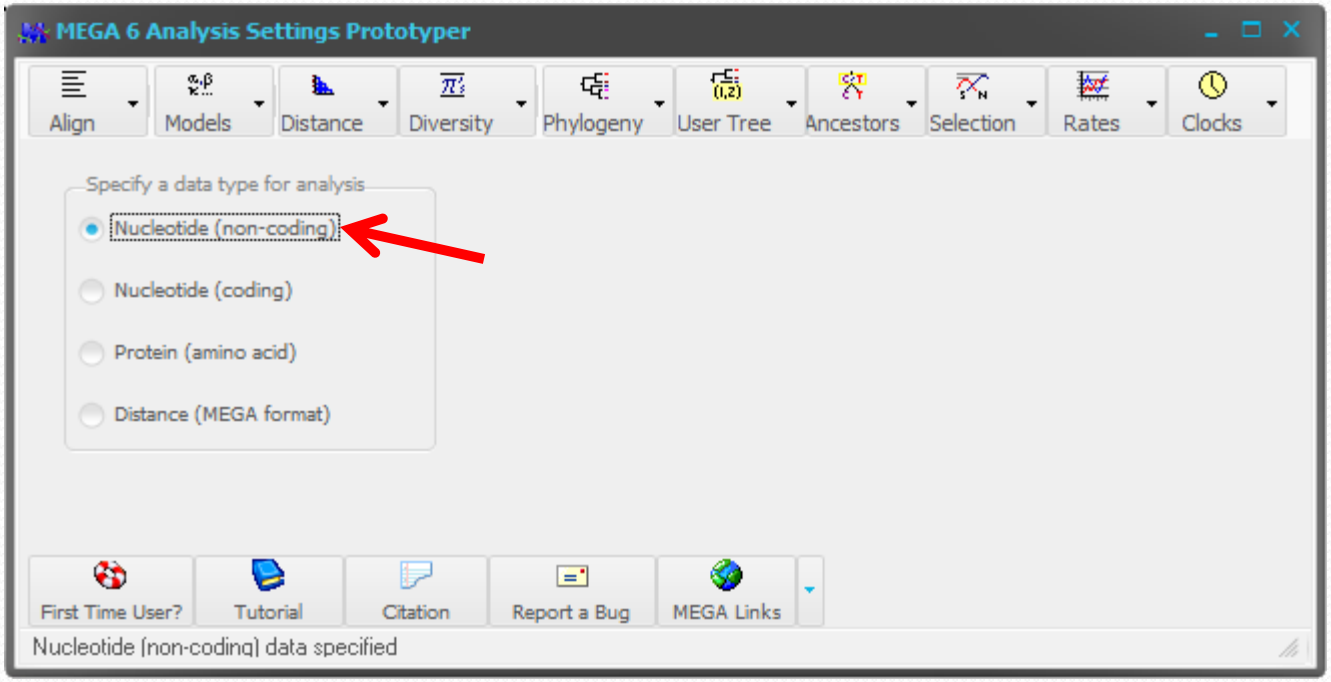
Step 1

- Open MEGA-Proto by selecting MEGA-Proto from the Start Menu



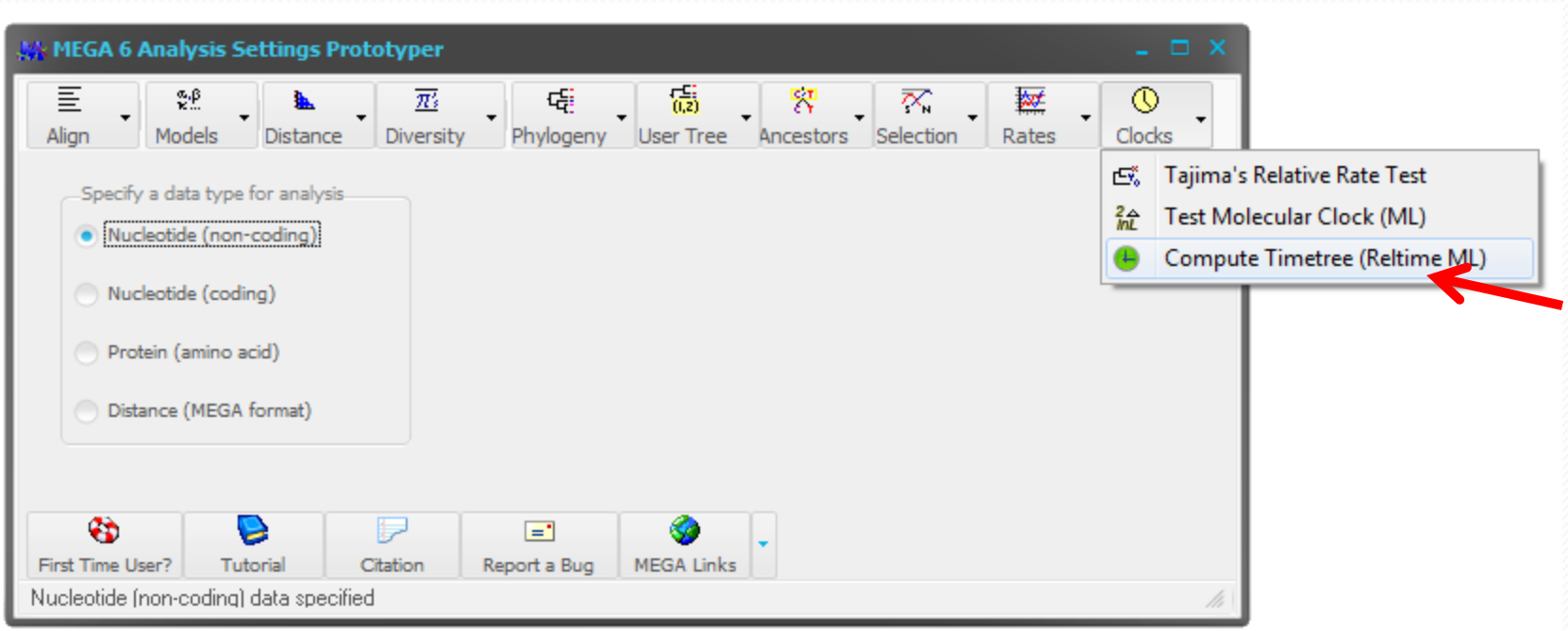
Step 2

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



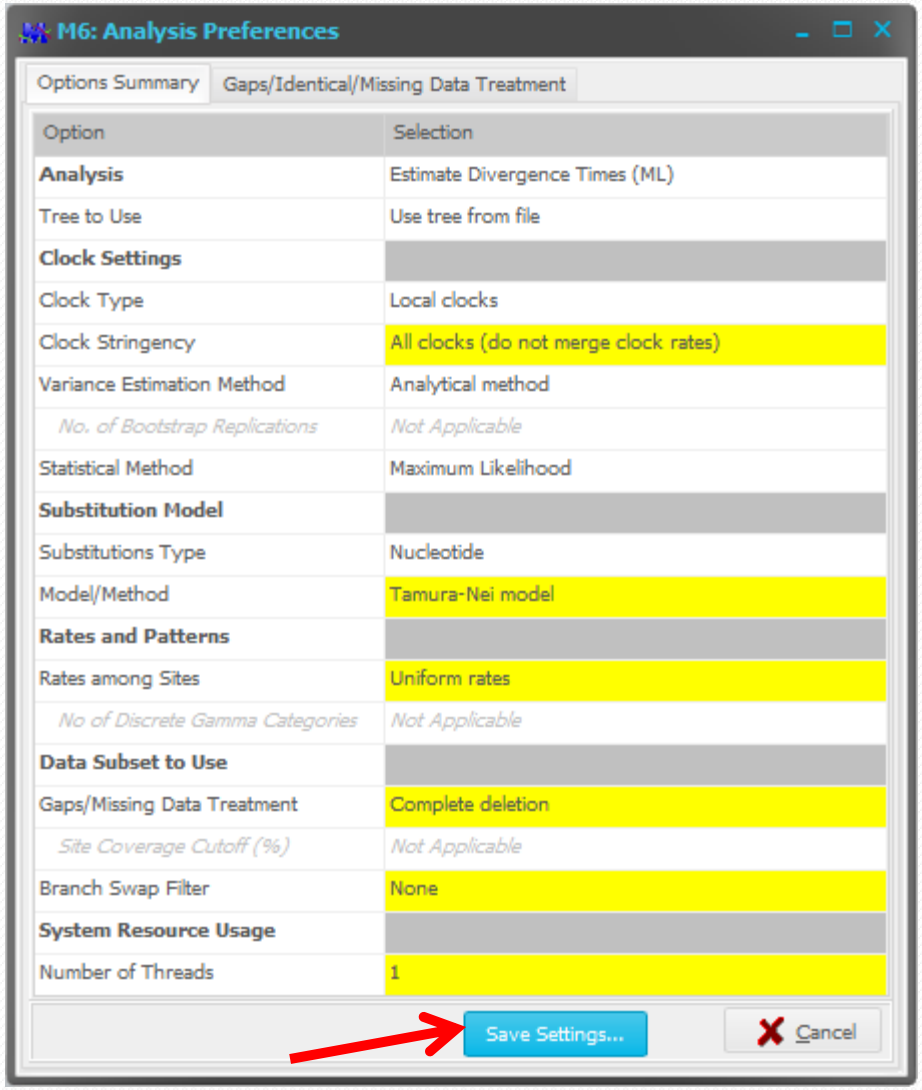
Step 3

- Select *Compute Timetree (Reltime ML)* from the *Clocks* menu.



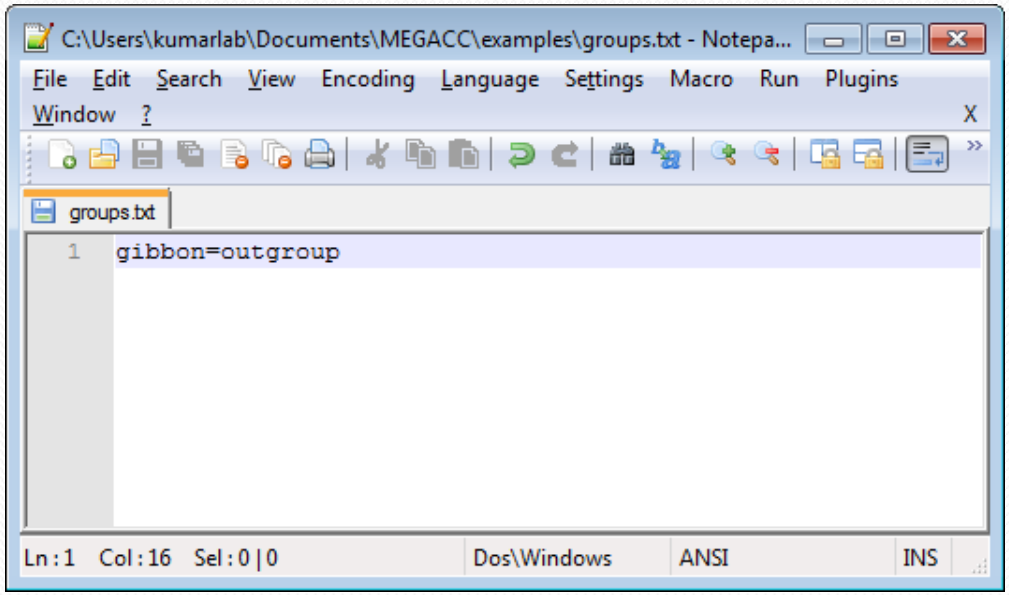
Step 4

- Adjust the analysis preferences to match those shown.
- Click the *Save Settings…* button and save the analysis options file as *demoSettings.mao* in the MEGACC\examples directory.



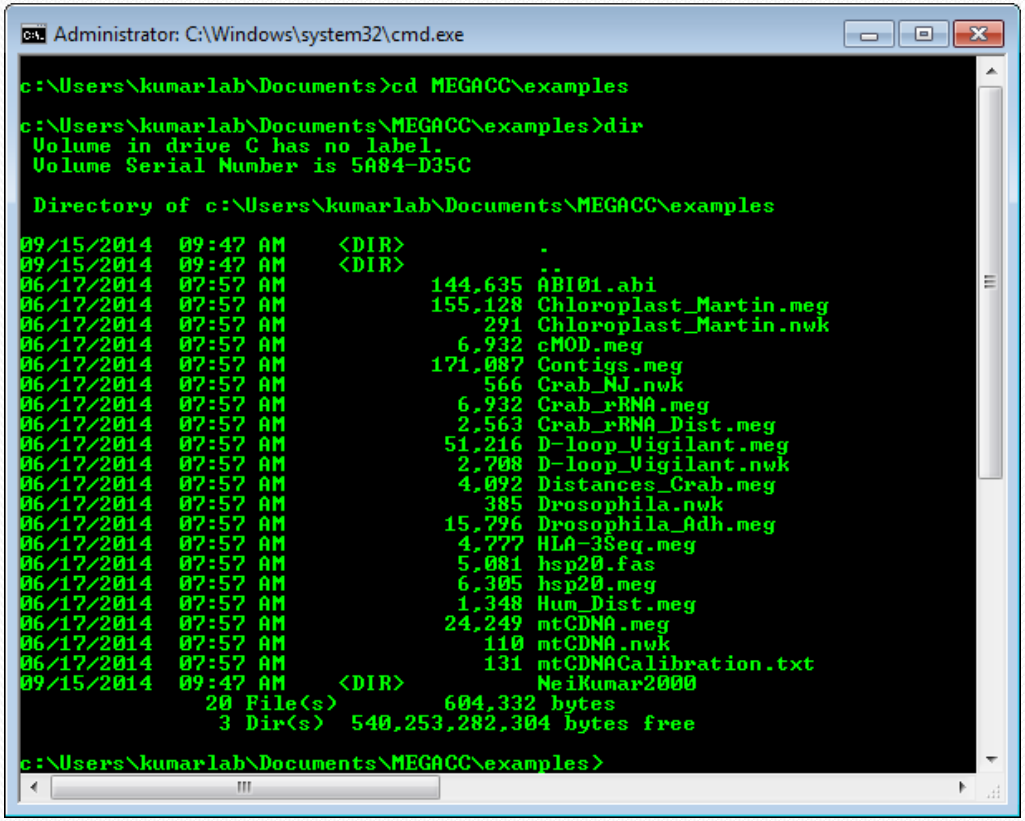
Step 5

- The timetree analysis requires that we specify an outgroup. To do so, create a text file and add the line ‘gibbon=outgroup’. Save this file as groups.txt in the MEGACC\examples directory.



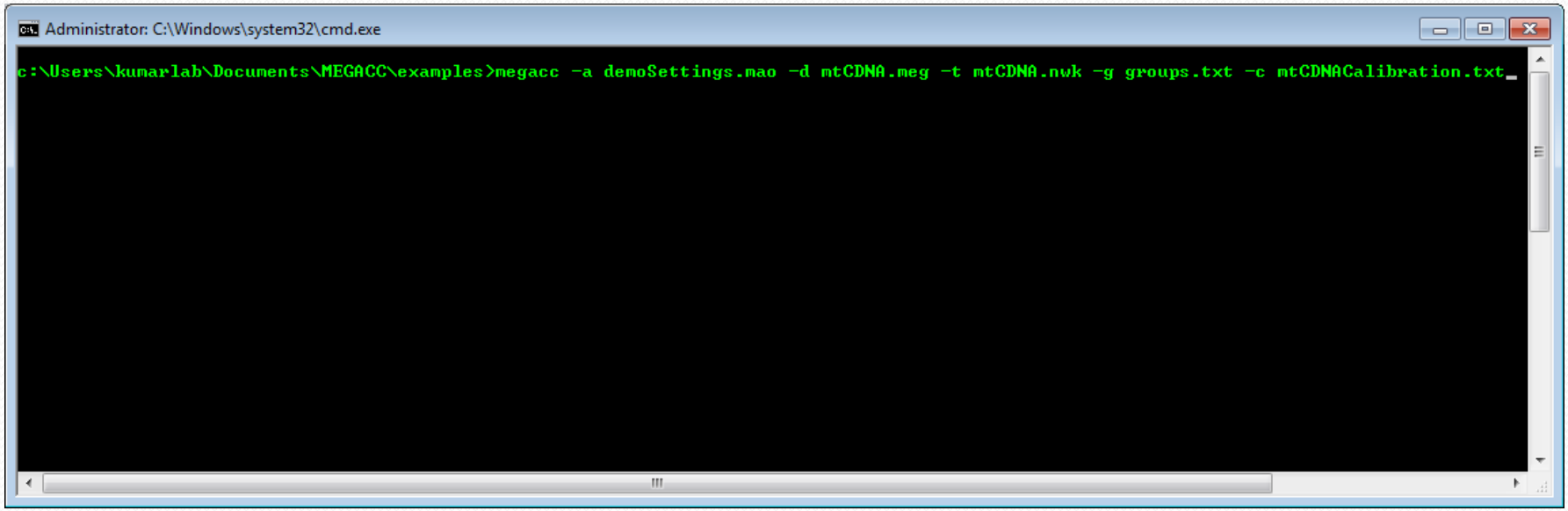
Step 6

- Open a command prompt.
- Navigate to the MEGACC\examples directory using the *cd* command



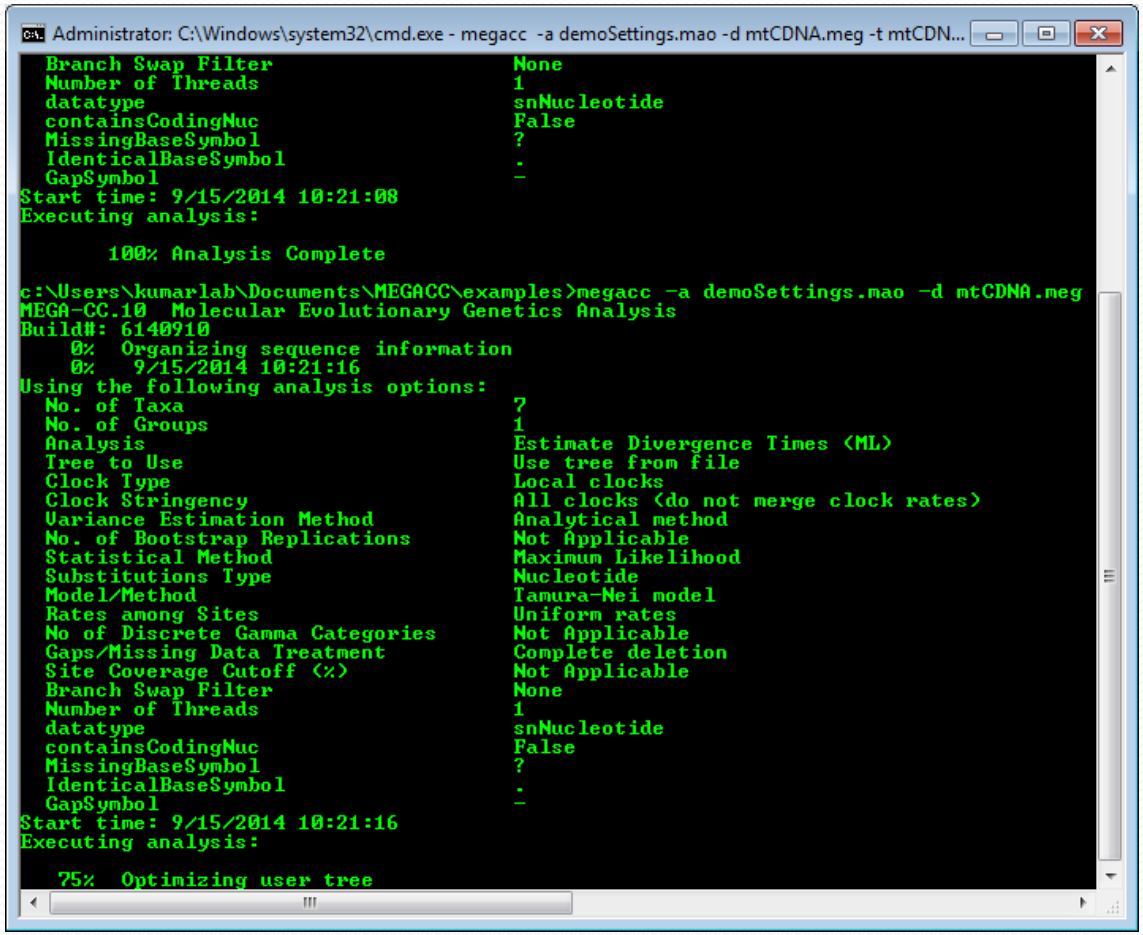
Step 7

- Execute the analysis by calling megacc from the command prompt as follows:
- megacc –a demoSettings.mao –d mtCDNA.meg –t mtCDNA.nwk –g groups.txt –c mtCDNACalibration.txt



Step 8

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

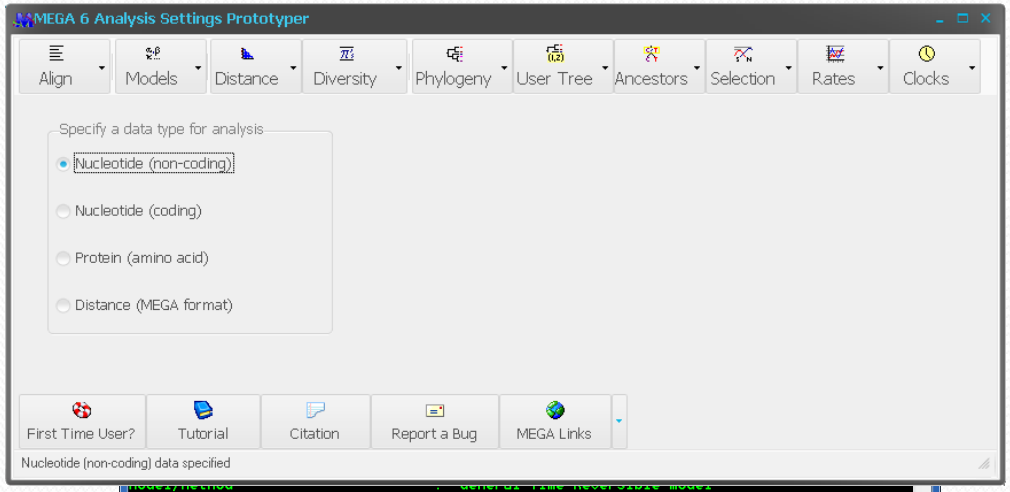


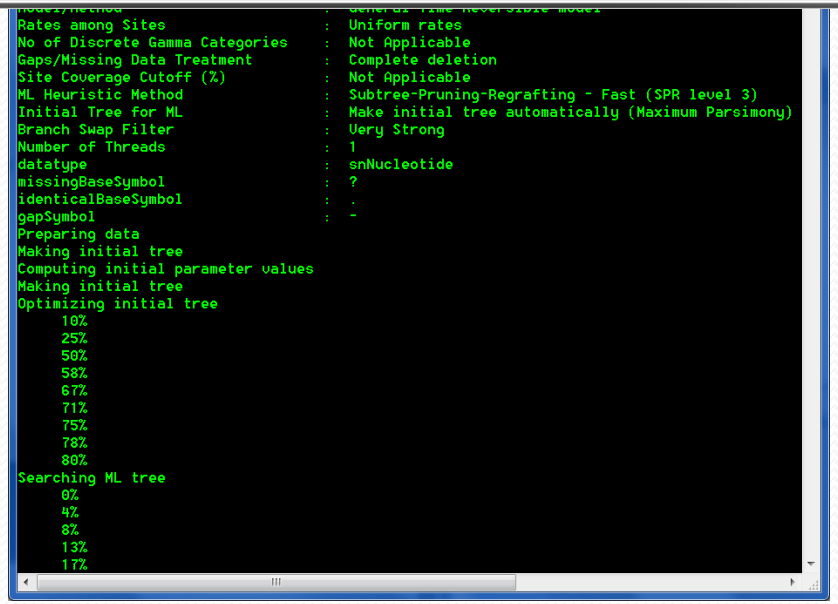
Finally

- The analysis will produce several output files in the directory MEGACC\examples\M6CC_Out
  - mtCDNA-xxxx_exactTimes.nwk
    - This Newick file gives the timetree scaled according to the estimated divergence times.
  - mtCDNA-xxxx_relTimes.nwk
    - This Newick file gives the timetree scaled according to the estimated relative divergence times.
  - mtCDNA-xxxx.txt
    - This text file gives a more detailed representation of the timetree, including relative times, exact times, evolutionary rates, and divergence time std errors.
  - mtCDNA-xxxx_summary.txt
    - This file gives analysis information such as the log likelihood value of the Maximum Likelihood tree, ts/tv ratio, etc…

Demo2

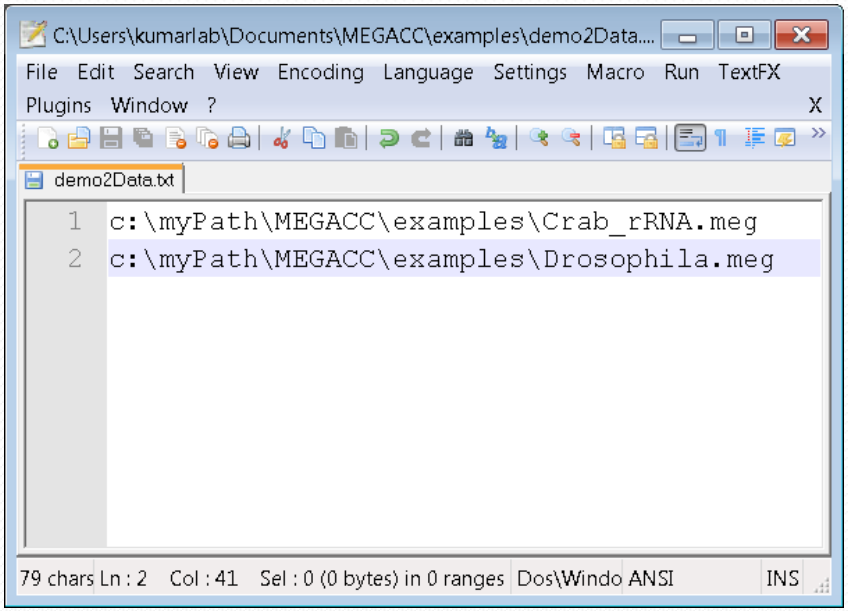
- The following example demonstrates how to use the File Iterator system in MEGA-CC to process multiple input data files using a single analysis options file.





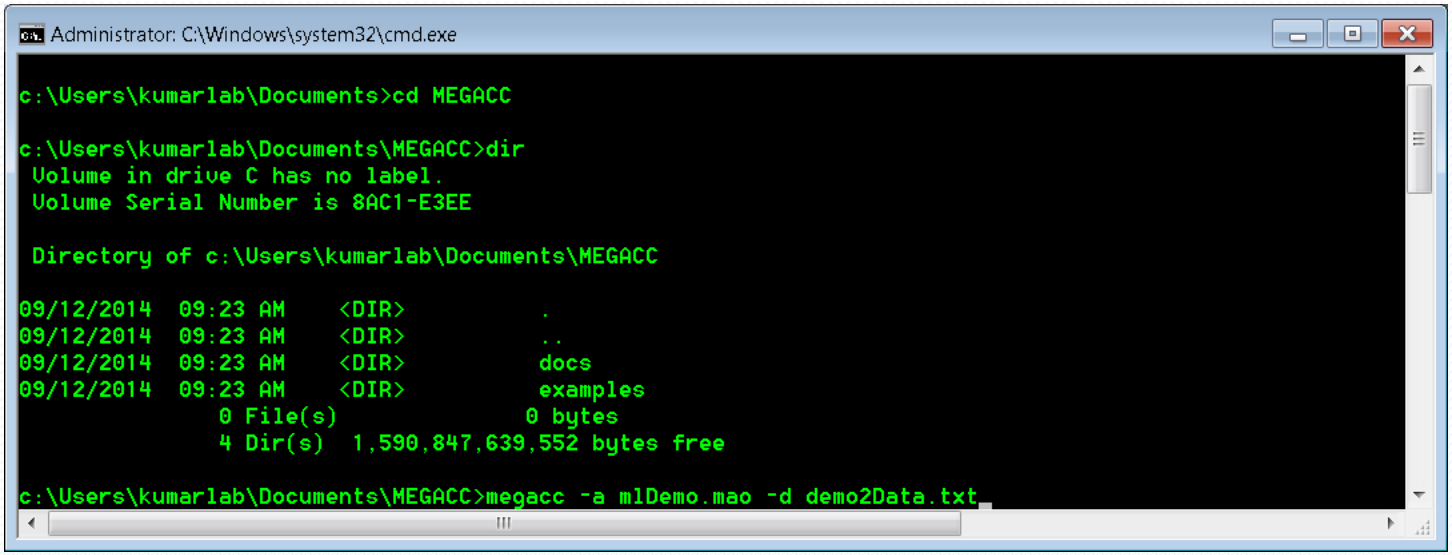
Step 1

- 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.



Step 2

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



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!