

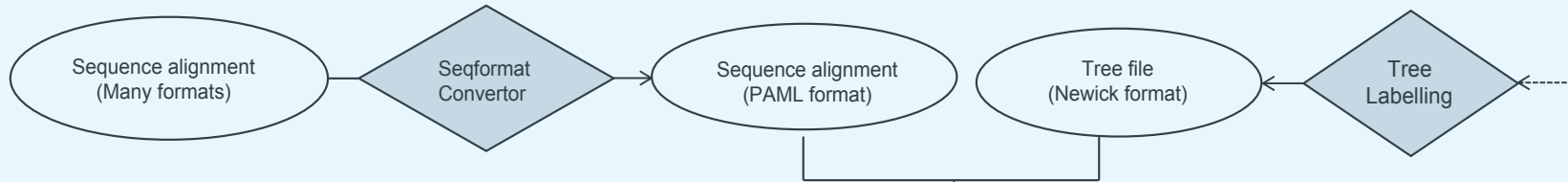
EasyCodeML: A visual tool for analysis of selection using CodeML

A quick guide to EasyCodeML

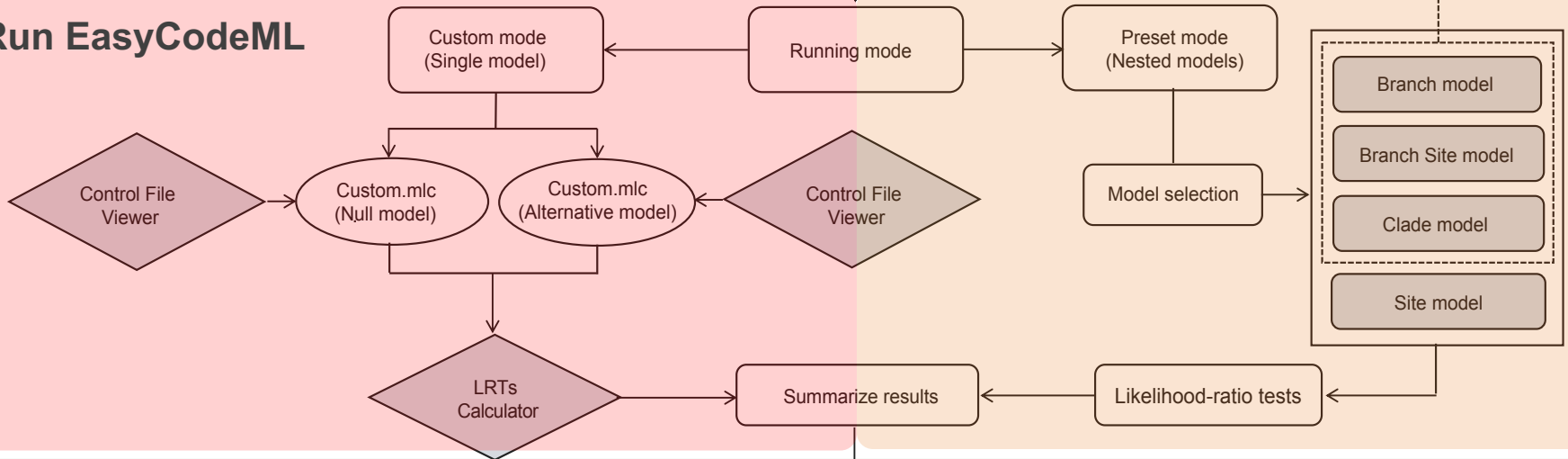
16 Jan, 2019
Fangluan Gao et. al

EasyCodeML workflow

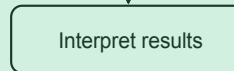
Prepare data



Run EasyCodeML



Interpret results



I Prepare data

- **Sequence alignment in PAML format**

EasyCodeML fully supports PAML format (e.g., Examples/Example1.pml). We have incorporated a file-format convertor ([Fig. 1](#)) that can convert diverse types of data (e.g., Clustal, FASTA, MEGA, Nexus, and Phylip formats) into PAML format.

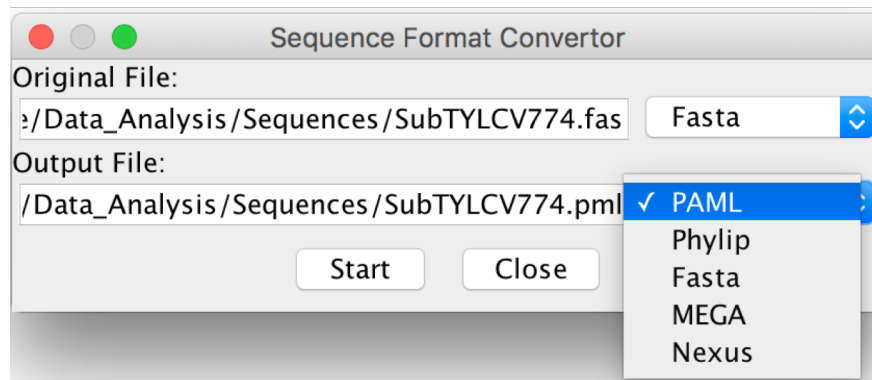


Fig. 1 Format conversion in Sequence Format Converter

I Prepare data

● Tree file in Newick format

The tree file must be in plain Newick format (e.g., Examples/Example1.tree; [Fig. 2](#)). Branch lengths can be included and each node can have a single label. Taxon names cannot have illegal characters (such as spaces and semicolons).

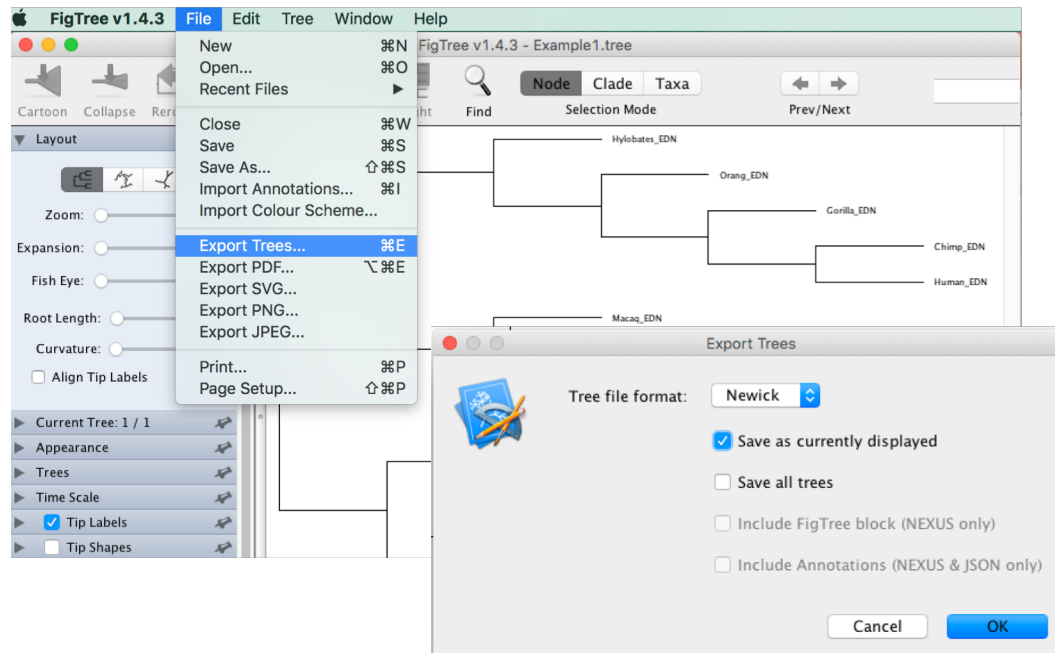


Fig. 2 Exporting a tree in Newick format



Prepare data

● Tree labelling

Tree labelling is necessary when using the branch-related models (branch models, branch-site models, and clade models; [Fig. 3](#)), but not with the site models.



Fig. 3 Labelling branches in a tree for the (A) branch and branch-site and (B) clade models

● **Preset running mode**

1. Select a local folder as working directory
2. Select a model type for selection analysis
3. Load sequence alignment in PAML format
4. Load tree file in Newick format
5. Check consistency of taxon labels between the tree and sequence files
6. Label branches or clades in the phylogenetic tree (branch-related models only)
7. Save parameter settings
8. Start CodeML analysis in normal /silent modes
9. Perform a likelihood-ratio test for the nested models
10. Export a publication-quality table that contains the results
11. Launch Microsoft Excel to view the saved results file

II Run EasyCodeML

The screenshot shows the EasyCodeML v1.2 application window. The interface is divided into several sections: Running Mode, Setup, and Likelihood Ratio Test (LRT). Numbered callouts (1-11) highlight specific elements:

- 1** Working Directory: sers/raindy/Documents/Biosoft/EasyCodeML/Example/
- 2** Model Selection: Branch Model, Branch Site Model, Site Model, Clade Model (selected)
- 3** Aligned Sequence File in PAML Format: syCodeML/Test0130/Example1.pml
- 4** Tree File in Newick Format: /Users/raindy/Documents/Biosoft/EasyCodeML/
- 5** Check button
- 6** Label button
- 7** Save Current Profile button
- 8** Run CodeML button
- 9** Test Statistic: $2 \Delta \ln L = 2 \times (-1686.673926 - -1688.777881) = 4.2079100000000835$
- 10** Export button
- 11** View button

Step 1
Load data and
configure parameters

Step 2
CodeML analysis

Step 3
Summarize results

II Run EasyCodeML

● Custom running mode

1. Select a local folder as working directory
2. Load sequence alignment in PAML format
3. Load tree file in Newick format
4. Choose a data type
5. Check consistency of taxon labels between the tree and sequence files
6. Label branches or clades in the phylogenetic tree (branch-related models only)
7. Load a codon model from Control File Viewer ([Fig. 4](#))
8. Save parameter settings
9. Start CodeML analysis
10. View the saved results file
11. Run a likelihood-ratio test ([Fig. 5](#))



Run EasyCodeML

The screenshot shows the EasyCodeML v1.2 application window. It has a menu bar with 'File', 'Tools', 'Options', and 'Help'. The interface is divided into several sections:

- Running Mode:** Two radio buttons, 'Preset (Nested Models)' and 'Custom (Single Model)'. The 'Custom' button is selected and circled with a blue circle containing the number 1.
- Setup:** A section containing file paths and other settings. It includes:
 - 'Working Directory' with a text box and a browse button (...).
 - 'Aligned Sequence File in PAML Format' with a text box and a browse button (...).
 - 'Tree File in Newick Format' with a text box and a browse button (...).
 - 'Num of Threads' with a spinner box set to 2.
 - 'Codons' with a dropdown menu.
 - 'Check' and 'Label' buttons.
- Parameters:** A large section with various dropdown menus and input fields for configuring the analysis, including Model, Fix κ , Fix ω , Fix α , Codon Frequency, AA Distance, AA Rate File, Small Difference, Run Mode, Clock, Mgene, ICode, and checkboxes for 'different alpha's for genes(Mapha)', 'getSE', 'RateAncestor', and 'Clean data'.
- Buttons:** A row of buttons at the bottom of the main panel: 'Load', 'Reset', 'Save Current Profile', 'Run CodeML' (highlighted with a blue circle containing the number 9), 'Stop CodeML', and 'View'.
- Running Status:** A section at the bottom with a 'Main' button and a text area containing a welcome message and instructions.

Numbered callouts are placed over the interface:

- 1: Custom (Single Model) button
- 2: Aligned Sequence File in PAML Format text box
- 3: Tree File in Newick Format text box
- 4: Codons dropdown menu
- 5: Check button
- 6: Label button
- 7: Load button
- 8: Save Current Profile button
- 9: Run CodeML button
- 10: View button

Step 1
Choose a running mode and input data

Step 2
Configure the running parameters

Step 3
CodeML analysis

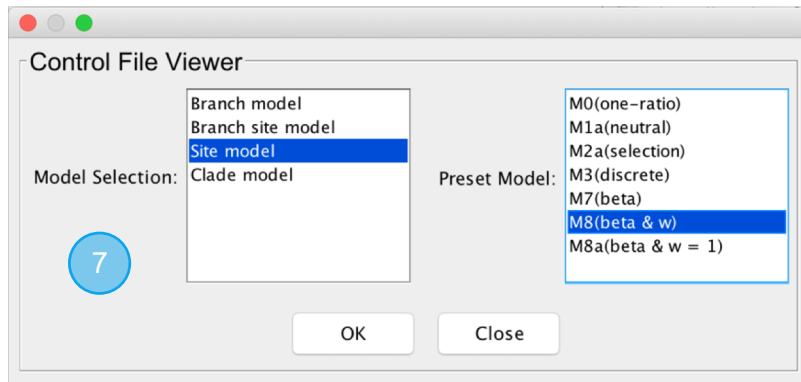


Fig. 4 Control File Viewer that includes codon models with pre-optimized parameters

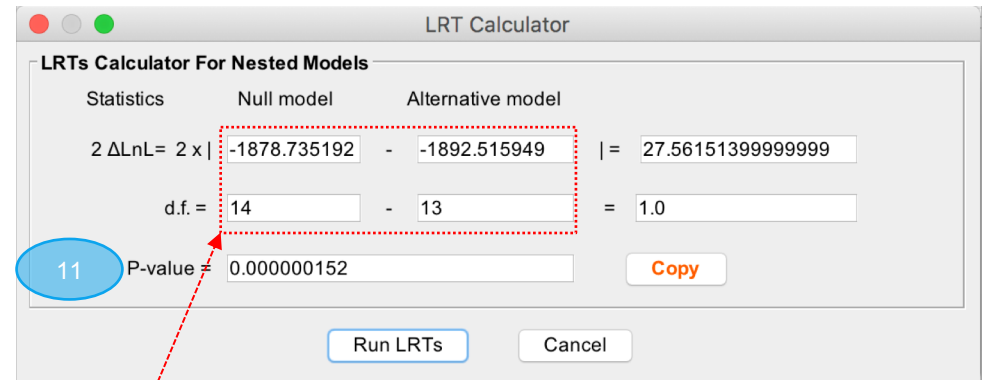


Fig. 5 LRT Calculator for running a likelihood-ratio test

M8a:

TREE # 1: (((4, 6), 2), 5), 1, 3); MP score: 58

lnL(ntime: 9 np: 13) -1892.515949 +0.000000

7..8 8..9 9..10 10..4 10..6 9..2 8..5 7..1 7..3

0.043343 0.011395 0.002754 0.005543 0.011176 0.000004 0.002428 0.040393 0.098672 1.288211 0.828982 0.005000

3.272293

M8:

TREE # 1: (((4, 6), 2), 5), 1, 3); MP score: 58

check convergence..

lnL(ntime: 9 np: 14) -1878.735192 +0.000000

7..8 8..9 9..10 10..4 10..6 9..2 8..5 7..1 7..3

0.057571 0.012156 0.002972 0.006652 0.013314 0.000004 0.003824 0.064314 0.163142 1.617384 0.966804 0.031342

0.166519 19.657673

Troubleshooting

- **Java 1.6 or higher**

EasyCodeML requires Java Runtime Environment (JRE) 1.6 or higher.

To check your version of Java, type **java -version** at a command prompt (CMD in Windows, terminal in Mac or Linux).

The latest version of Java can be downloaded from <http://www.java.com>

- **Run EasyCodeML in debug mode**

To run EasyCodeML in debug mode, type **java -jar EasyCodeML.jar** at the command prompt (CMD in windows or Terminal in Mac or Linux).

Contact us

- **EasyCodeML is an ongoing project**

We welcome bug reports, feedback, and suggestions

- <https://github.com/BioEasy/EasyCodeML>

- **Contact us**

Fangluan Gao	raindy@fafu.edu.cn
Chengjie Chen	ccj0410@gmail.com