

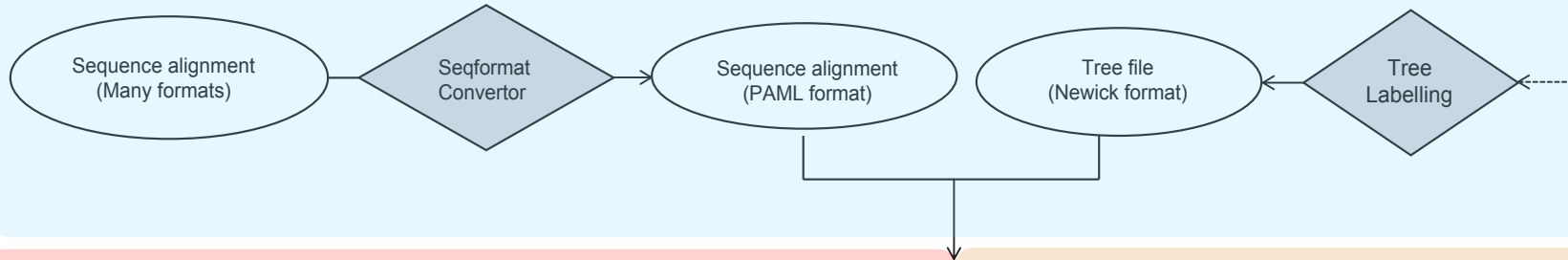
EasyCodeML: A visual tool for analysis of selection using CodeML

A quick guide to EasyCodeML

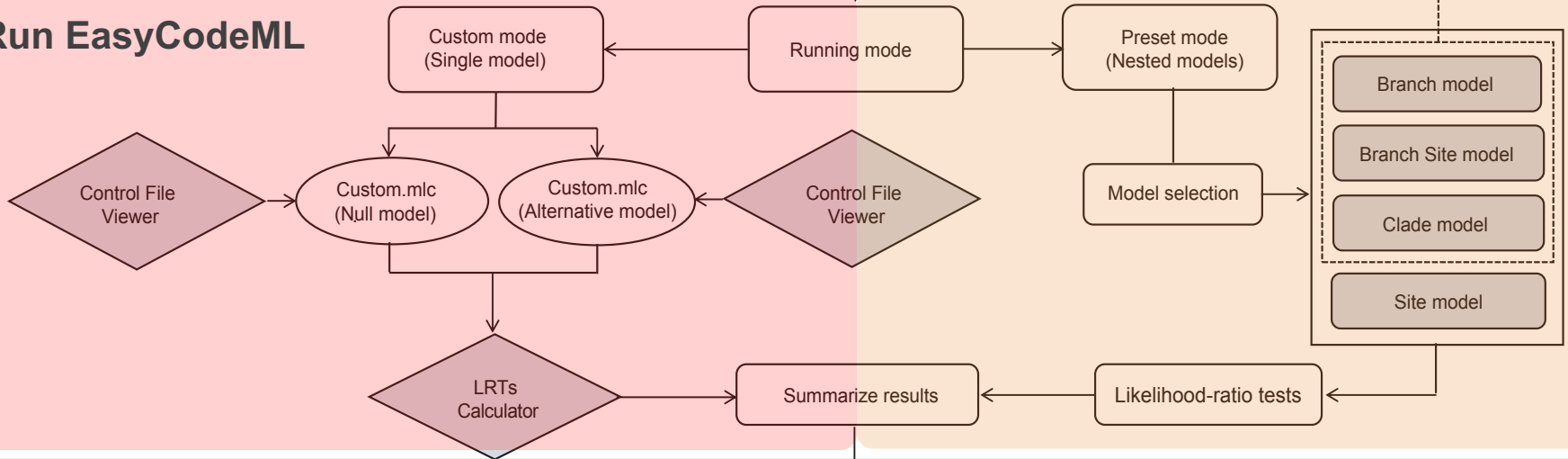
3 Oct, 2018
Fangluan Gao et. al

EasyCodeML workflow

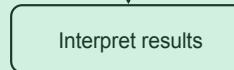
Prepare data



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.

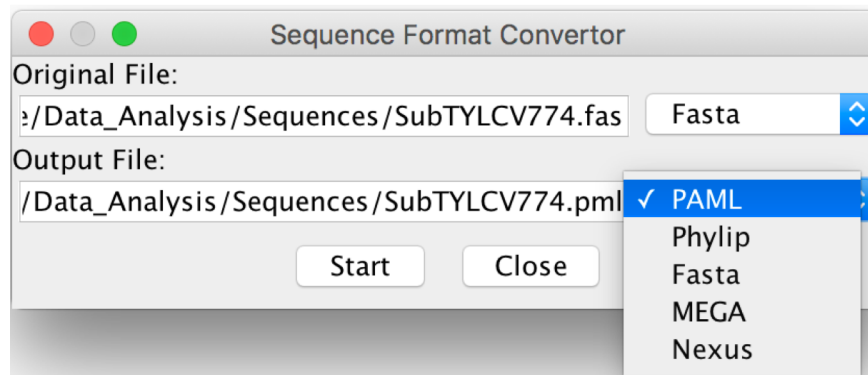


Fig. 1 Format conversion in Sequence Format Converter

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).

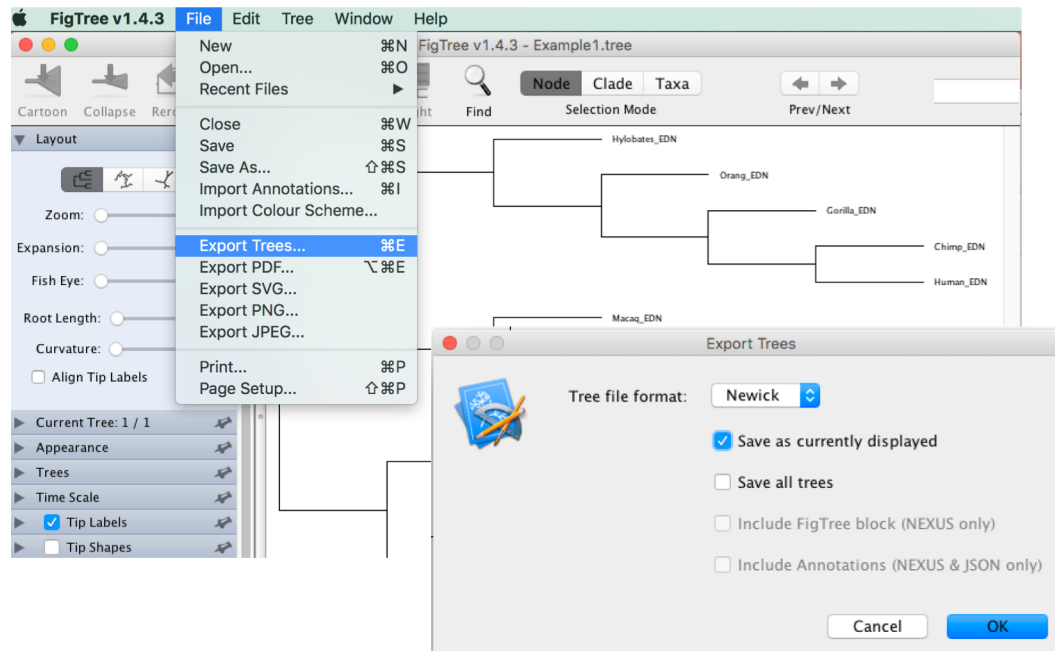


Fig. 2 Exporting a tree in Newick format



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

II Run EasyCodeML

● **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
8. Perform a likelihood-ratio test for the nested models
9. Export a publication-quality table that contains the results
10. Launch Microsoft Excel to view the saved results file

II Run EasyCodeML

The screenshot shows the EasyCodeML v1.1 application window. The interface is divided into several sections: 'Running Mode', 'Setup', 'Likelihood Ratio Test (LRT)', and a results section at the bottom. Numbered blue circles (1-11) are overlaid on the interface to indicate the sequence of steps. Step 1 is in the 'Running Mode' section, Step 2 is in the 'Setup' section, and Step 3 is in the 'Likelihood Ratio Test (LRT)' section. The 'Running Mode' section has two radio buttons: 'Preset (Nested Models)' (selected) and 'Custom (Single Model)'. The 'Setup' section contains fields for 'Working Directory', 'Model Selection' (with radio buttons for 'Branch Model', 'Branch Site Model', 'Site Model', and 'Clade Model' which is selected), 'Aligned Sequence File in PAML Format', 'Tree File in Newick Format', 'Num of Threads' (set to 2), and a 'Clean data' checkbox (checked). There are 'Check' and 'Label' buttons. The 'Likelihood Ratio Test (LRT)' section shows the 'Current Model' as 'Clade Model (M2a_rel vs. CmC)'. It displays the 'Test Statistic' as $2 \Delta \ln L = 2 \times | -1686.673926 - (-1688.777881) | = 4.2079100000000835$, with 'd.f.' calculated as $33 - 32 = 1$ and a 'P-value' of 0.040235883. The bottom section shows the 'Summary of Results' file path and 'Export' and 'View' buttons.

EasyCodeML v1.1

File Tools Options Help

Running Mode

☒ **Preset** (Nested Models) ☐ **Custom** (Single Model)

Setup

1 Working Directory: ... | Num of Threads 2

2 Model Selection: ☐ Branch Model ☐ Branch Site Model ☐ Site Model ☒ Clade Model

3 Aligned Sequence File in PAML Format: ... ☒ Clean data

4 Tree File in Newick Format: ... 5 Check 6 Label

7 Save Current Profile Stop Current Running 8 Run CodeML Run LRTs

Likelihood Ratio Test (LRT)

Current Model: Clade Model (M2a_rel vs. CmC)

9 Test Statistic: $2 \Delta \ln L = 2 \times | -1686.673926 - (-1688.777881) | = 4.2079100000000835$

d.f. = 33 - 32 = 1 P-value = 0.040235883

Summary of Results: ... 10 Export | View 11

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 v1.1 application window. The interface is divided into several sections:
1. **Running Mode**: Two radio buttons, 'Preset (Nested Models)' and 'Custom (Single Model)'.
2. **Setup**: Fields for 'Working Directory', 'Aligned Sequence File in PAML Format', and 'Tree File in Newick Format'. It also includes 'Num of Threads' and a dropdown for '1:codons'.
3. **Parameters**: A large section with various settings like '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)', 'RateAncestor', and 'Clean data'.
4. **Buttons**: A row of buttons including 'Load', 'Reset', 'Save Current Profile', 'Run CodeML' (highlighted in orange), 'Stop CodeML', and 'View'.
5. **Running Status**: A text area at the bottom with a 'Main' button and a scrollable log of messages.

Numbered callouts are placed as follows:
1. Working Directory field
2. Aligned Sequence File field
3. Tree File field
4. 1:codons dropdown
5. Check button
6. Label field
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

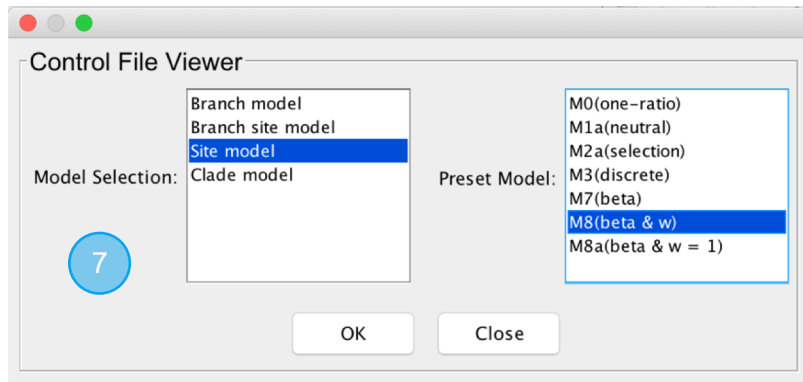


Fig. 4 Control File Viewer that includes codon models with pre-optimized parameters

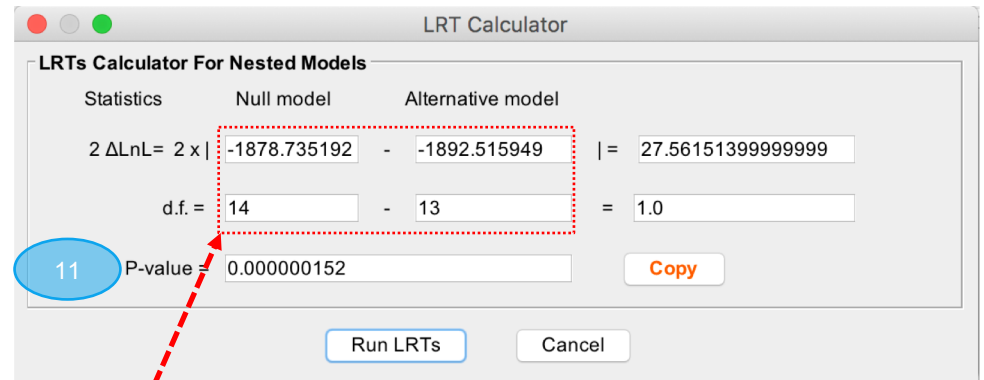


Fig. 5 LRT Calculator for running a likelihood-ratio test

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