

## WORKED EXAMPLE 2

We briefly illustrate the use of the custom running mode in EasyCodeML by using a data set from Padhi et al. (2009). We compare the M8 and M8a models to test for sites under positive selection in the outer membrane protein C (*ompC*) of strains of *Enterobacter aerogenes*. This particular model comparison is not available in the preset mode of EasyCodeML.

### 1. Choosing a running mode and input data

We switch current running mode to the custom mode and specify a local folder as the working directory using drag-and-drop, as described above for the preset mode.

### 2. Configure the running parameters

The “Load” button can be used to load a codon model available from a control file viewer (Fig. 5A). This will bring up a dialog box from which we choose the M8a model. We can further modify the various parameter values to meet different requirements. Tree labeling is necessary when examining the branch-related models (branch models, branch-site models, and clade models), but not with the site models. Therefore, default values are used for all parameters except for leaving “Clean data” unchecked (Fig. 2B). We need to save the current profile by clicking the “Save Current Profile” button after checking if the taxon labels match between the tree and sequence files.

### 3. CodeML analysis

Clicking “Run CodeML” will start the analysis. In order to perform the subsequent likelihood-ratio test, we will need to run both models. Therefore, we need to repeat the procedure for the M8 model.

We navigate to the working directory and locate the main result files (mlc) of the model M8 and M8a. After noting the log-likelihood (lnL) values and the number of parameters (np) in these mlc files, we enter them in the LRT calculator from the ‘Tools’ menu and run a likelihood-ratio test. Based on the lnL and np values of the null model (M8, lnL = -1878.7, np=14) and the alternative model (M8a, lnL = -1892.5, np=13), the test yields a p-value below 0.05 (Fig. 4A).

### 4. Identification of sites under selection

In the comparison of models M8 and M8a, the BEB analysis under model M8 is used to identify codons under positive selection. Thus, we find a block called “Bayes Empirical Bayes (BEB) analysis” in the mlc file. This block lists the amino acids that have a BEB score higher than 0.5. Sites potentially under positive selection are suggested by BEB values higher than 95%, which are indicated by asterisks. In this data set, we identified nine codons as being under positive selection with posterior probability >0.95, matching the results of Padhi et al. (2009).