Phylogenetic theory and practice

Stanford Earth Skillshare

Hanon Mcshea

Nov 8, 2019

Slides and all lab materials are at https://www.github.com/cyclase/skillshare

Setup

1. Make a free account on the CIPRES ("cypress") Science Gateway at https://www.phylo.org/portal2/register.action.
2. Create New Folder for yourself on CIPRES.
3. Download datafiles (DSRa_spnames.translated .nex and .phy) from github.
4. Click on the Data section of your CIPRES folder and upload the 2 datafiles.

Evolutionary model testing

1. Click on the **Tasks** section of your CIPRES folder and the **Create New Task** button. Type "model" in the Description box and select the .phy file as the input and ModelTest-NG as the tool.
2. Check The dataset contains amino acid (-d aa).
3. Set select the candidate model's rate heterogeneity (-h) to "Both +I and +G"
4. Click **Advanced Parameters** and set the seed value to 12345.
5. Scroll down and click **Save Parameters** at the bottom and **Ok** on the little dialog bubble.
6. Click **Save and Run Task** and **Ok** again.
7. On the Tasks list click the View Output button, and View the STDOUT file. What model does the program recommend? What is the proportion of invariant sites? What is the gamma shape parameter?

Maximum likelihood tree estimation

1. Click on the **Tasks** section of your CIPRES folder and the **Create New Task** button. Type "ML" in the Description box and select the .phy file as the input and IQ-Tree as the tool.
2. Click the parameters button. How many partitions does your data set have? 1
3. Select the kind of run you need: Tree Inference
4. Specify your sequence type (-st): Amino Acid
5. Specify the seed (-seed): 12345
6. Click **Advanced Parameters** and scroll down to Specify Substitution Models. Specify a Protein model (-m) as the best model returned by ModelTest.
7. Specify rate heterogeneity as "Invariable site plus discrete Gamma model".
8. Specify the Gamma shape parameter (-a) and the proportion of invariant sites (-i) as the output from ModelTest.

9. Scroll down to Bootstrap Search Parameters and set Specify number of bootstrap replicates to 1000.
10. Click **Save Parameters** at the bottom and **Ok** on the little dialog bubble.
11. Click **Save and Run Task** and **Ok** again.
12. Look at the STDOUT file for the run. What is the likelihood of the tree?
13. Go back to the output files and download output.treefile.

Bayesian tree estimation

1. Click on the **Tasks** section of your CIPRES folder and the **Create New Task** button. Type "Bayesian" in the Description box and select the **.nex** file as the input and MrBayes on XSEDE as the tool.
2. My Data Type Is protein.
3. Set the seed and swapseed numbers to 12345.
4. Under Likelihood Model Parameters, set the number of substitution types (Nst=) to mixed.
5. Set the nucleotide substitution model (Nucmodel=) to protein.
6. Set the model for among-site rate variation (Rates=) to invgamma.
7. Under Configure Priors, Set the Rate Matrix for Amino Acids (Aamodelpr=) to the best model returned by ModelTest.
8. Under Parameters for MCMC, set Number of Generations (Ngen=) to 1000000 (one million).
9. Set How often should the Markov chain be sampled to 1000.
10. Under Set Sumt parameters, set Choose the output format for your consensus tree (Conformat=) to Simple.
11. Scroll down and click **Save Parameters** at the bottom and **Ok** on the little dialog bubble.
12. Click **Save and Run Task** and **Ok** again.
13. Look at the STDOUT file. Did the analysis converge?
14. Download infile.nex.con.tre.

Results

1. Go to https://itol.embl.de, click **Upload**, and upload the maximum likelihood tree file. Do the same in another tab with the Bayesian one.
2. Click the **Advanced** tab and **Display** Bootstraps/metadata as text. What are the support values on each tree telling you?