Name:

Comparative Genomics

Phylogenetics Lab

IQ-TREE Example Data

1. How many species (sequences) were in this alignment?
2. What is the length (in nucleotides) of this alignment?
3. What is the number of invariant (constant or ambiguous constant) sites?
4. What is the proportion of parsimony-informative sites?
5. What do you think "parsimony-informative" means?
6. Which model was selected as the best fit, and according to what criteria?

Use this substitution model cheat sheet (http://www.iqtree.org/doc/Substitution-Models) and the iq-tree output to answer the following questions:

1. How many degrees of freedom (df) does the selected model have?
2. Did the ModelFinder select the most complex model?
3. How does the chosen model handle base frequencies?
4. Why are some orthologs 'validated' and some 'predicted'? Think about our Genome Annotation module.
5. What is "bootstrap support", and how are those values calculated (hint: phylogenetics lecture)
6. How do you interpret bootstrap support values that are less than 100%?
7. Which branches have less than 100% bootstrap support?

1. What does that tell you about what we can infer about the evolution of the clades stemming from those branches? Use an example. For instance, if you want to describe a clade where A and B are closer to C, you can use parantheses: ((A,B) C). This describes a monophyletic A and B with a C outgroup.

The consensus tree summarizes all the bootstrap replicate trees.

Note where it indicates “Robinson-Foulds distance between ML tree and consensus tree: 0”?

1. What does this tell you?

BRCA1 Evolution

1. How many species (sequences) were in this alignment?
2. What is the length (in nucleotides) of this alignment?
3. What is the number of invariant (constant or ambiguous constant) sites?
4. What is the proportion of parsimony-informative sites?
5. What do you think "parsimony-informative" means?
6. Which model was selected as the best fit, and according to what criteria?

Use this substitution model cheat sheet (http://www.iqtree.org/doc/Substitution-Models) and the iq-tree output to answer the following questions:

1. How many degrees of freedom (df) does the selected model have?
2. Did the ModelFinder select the most complex model?
3. How does the chosen model handle base frequencies?
4. Which branches have less than 100% bootstrap support?
5. What does that tell you about what we can infer about the evolution of the clades stemming from those branches? Use an example.