Robert Petit

IBS594 – Phylogenetics Exercise 2

10/7/2014

1) Write a Methods and Results Section for your work. Include three figures (one for each tree). Make sure you include a caption for each figure, following conventions in published work on what information should be included in the caption. Make sure to include details of what programs you used, what model of evolution you used, how many generations you ran for your Bayesian analysis, etc.

**Methods**

The data used in these analyses were ITS DNA sequences from 12 species. These species include (GenBank accession in parenthesis) *Escovopsioides nivea* CBS 135749 (JQ815078), *Ascomycota sp*. RS054 (EU082786), *Escovopsis lentecrescens* CBS 135750 (JQ815079), *Escovopsis aspergilloides* CBS 423.93 (KF293287), *Escovopsis moelleri* CBS 135748 (JQ815077), *Escovopsis sp.* TF2CWG (FJ948131), *Escovopsis microspora* CBS 135751 (JQ815076), *Escovopsis weberi* ATCC 64542 (KF293285), *Escovopsis weberi* CBS 810.71 (KF293286), *Cladobotryum asterophorum* (EU340835), *Trichoderma viride* (EF568085), and an unknown Escovopsis/Escovopsioides species.

Using the default settings (GAPOPEN=15, GAPEXT=6.66) the ITS sequences were aligned using ClustalW (v2.1). The alignment from ClustalW was then converted from FASTA to Phylip using the shell script convertFasta2Phylip.sh available through RAxML. We predicted a model of evolution for our sequences using jModelTest (v2.1.6). jModelTest was run with the following parameters: -g 4 -i -f -AIC –a. The model with the top score based on Akaike Information Criterion (AIC ) was chosen for downstream maximum-likelihood and Bayesian analysis.

A parsimony tree and maximum likelihood tree was constructed using RAxML (rev. 051fd11). First a parsimony tree was constructed using the default settings of RAxML. We then constructed a maximum-likelihood tree using the parsimony tree as the starting tree and a GTRGAMMAI as our model of selection. The maximum-likelihood tree was bootstrapped 200 times before selecting the best tree. A bayesian tree was constructed using MrBayes (v3.2.2). MrBayes was run for 1 million generations with a burn-in of 25% using a GTRGAMMA model of selection. FigTree (v1.4.2) was used to visualize and annotate each of the three trees. Each of the trees were rooted between *Cladobotryum asterophorum* and *Trichoderma viride*. Also the known *Escovpsis* species were highlighted with green and known  *Escovopsioides*species were highlighted with blue. A Git repository of these data and methods is available at <https://github.com/rpetit3/ibs594-phylogenetics>.

**Results**

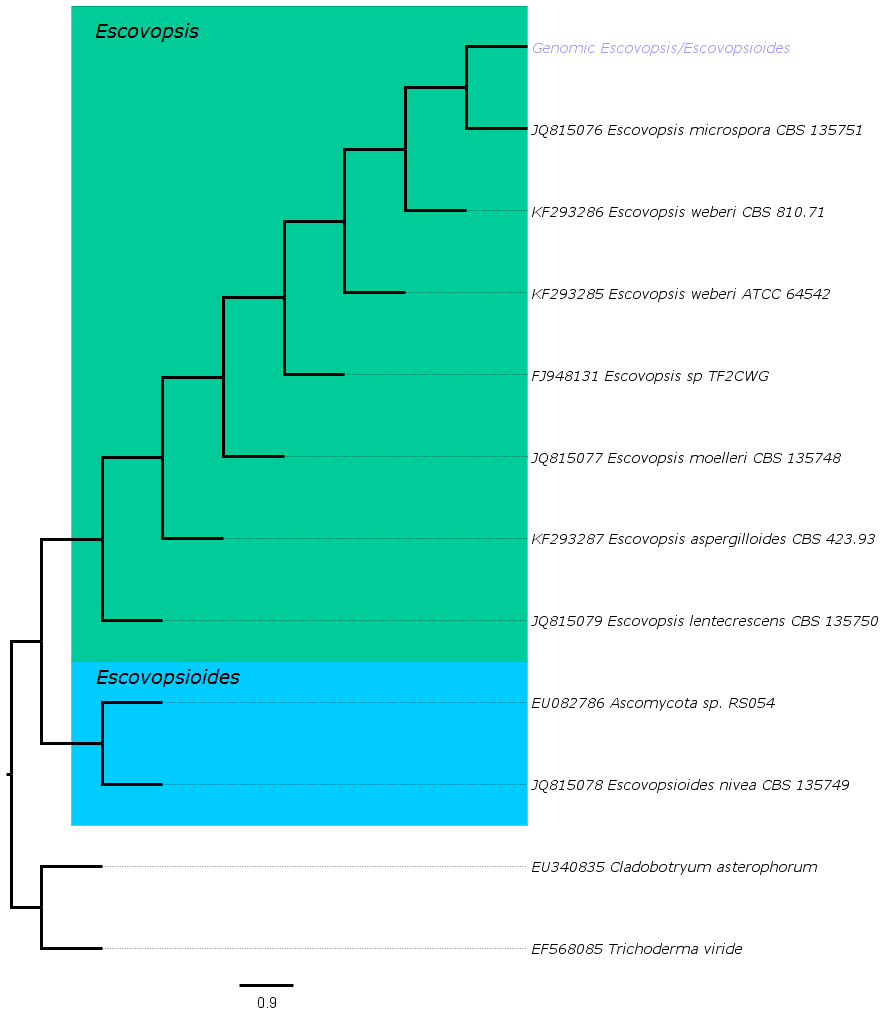
Since the splitting of the *Escovopsis* genus in to *Escovopsis* and *Escovopioides* there are still a species that must be placed into the proper genus. We set out to take an ITS sequence from the former *Escovopsis* genus and determine the proper genus (*Escovopsis* or *Escovopioides*). We compared our unknown ITS sequence against 11 other species of both *Escovopsis* and *Escovopioide*, also a few more distant ancestors.

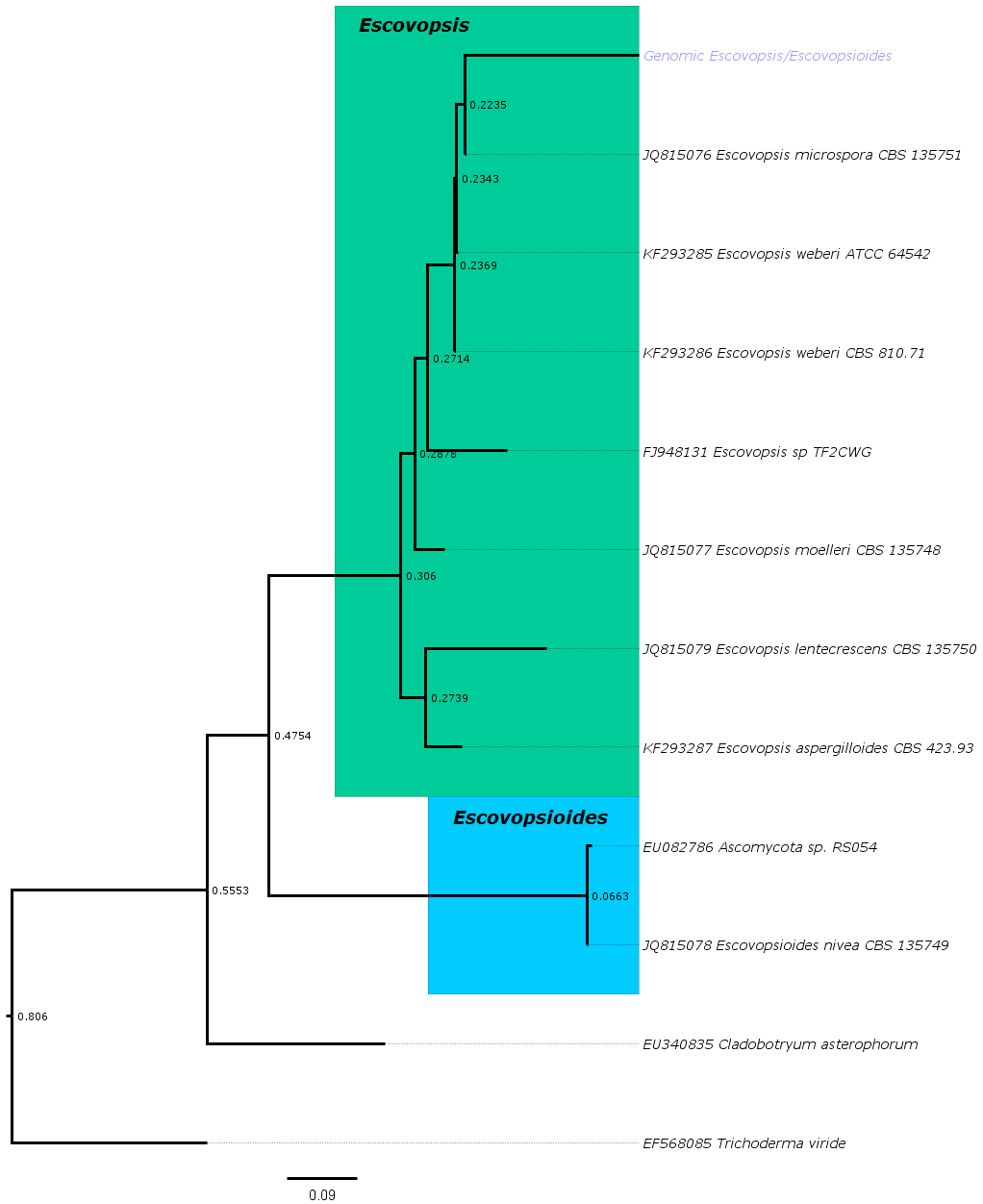
We used three phylogenetic methods to compare our sequences. Based on parsimony (Fig. 1), there is evidence that our unknown sample is likely of the genus *Escovopsis*. This trend continues with the maximum-likelihood tree (Fig. 2) and the Bayesian tree (Fig. 3) as well. In each of the three trees there is a clear monophyletic clade between genus *Escovopsis* and genus *Escovopioides*. Each method also produces evidence that our unknown sample is closely related to *Escovopsis microspora*.

It should be pointed out that our bootstrap values for both the maximum-likelihood and Bayesian trees are rather low. Internal nodes in both of these trees are less than 0.3, much less than the general rule of greater than 0.7 for strong confidence. Although these values do not necessarily mean the trees are incorrect, it does however affect the confidence in them. Most likely this is due to the nature of our data set. We have based these results on only 12 ITS sequences. Extending this analysis to include more genomic regions or accompanying morphology data could improve our confidence in the results.

2) In looking at your sequence alignment, how confident are you in each of your characters being homologous? Is this a problem? If so, how would you suggest someone using these data for a paper proceed?

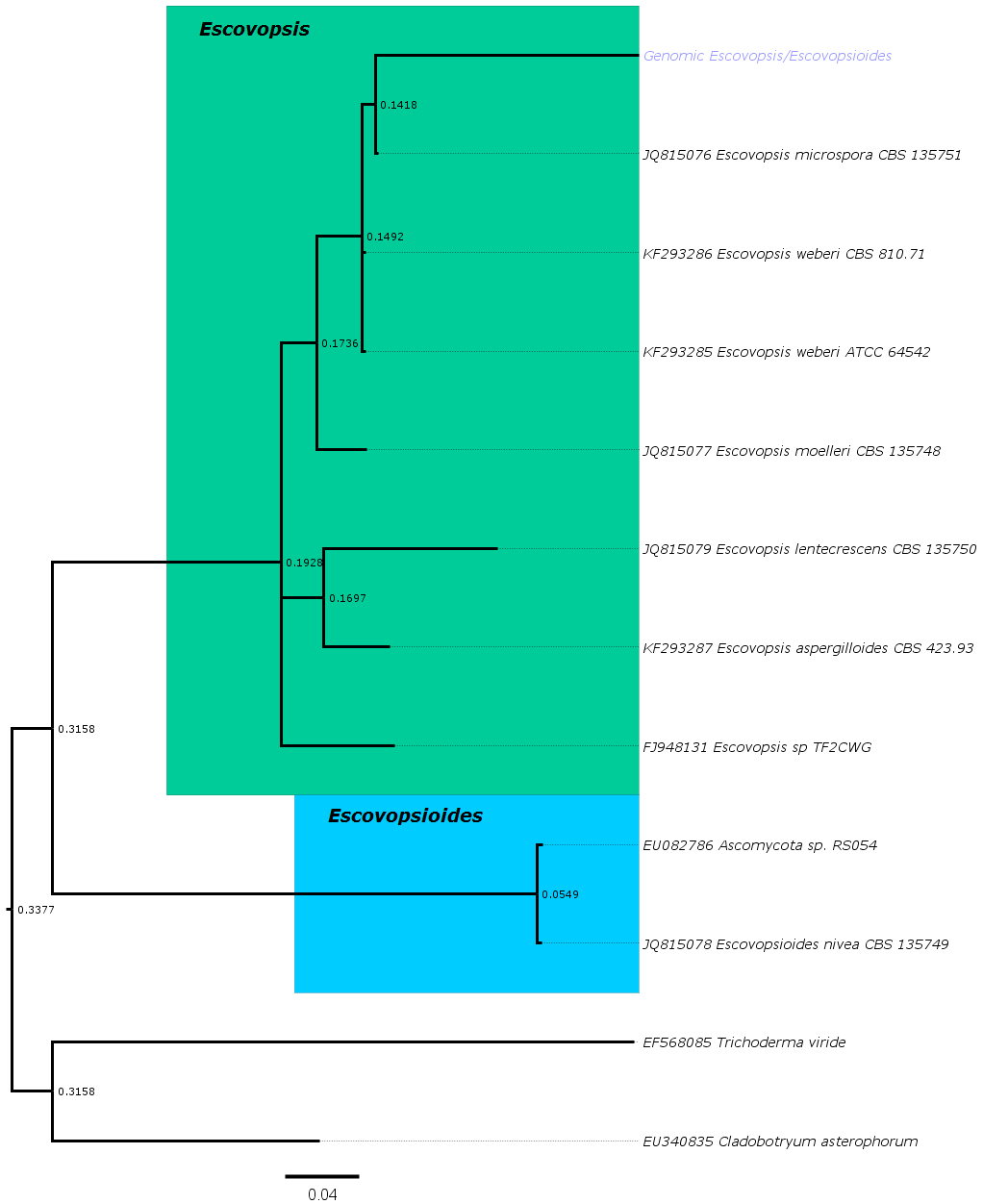
Looking at the alignments there are definitely some more ‘gappy’ (for lack of a better term) regions. There is a central area that is conserved among the sequences. I am confident in the alignments, but I would not base a paper solely on the analysis of these alignments. Surely you could go on a fishing expedition trying to achieve the best alignment by playing with the parameters and even trying different programs. You would probably even get better trees to base the results on. But, at the end of the day, what these results need is supporting evidence. This could be looking at morphological data: What morphological characteristics make our unknown sample more like *Escovopsis* or *Escovopsioide*? Are there morphological characteristics shared between our unknown sample and *Escovopsis microspore*? If the goal was to keep this study sequence driven, extend it to more genomic regions to gain a better resolution. I don’t know much about fungi, but there must be more than ITS that is conserved between the species. I imagine by extending the sequencing to more genomic regions the phylogenies will give much better support.

**Figure 1: Optimal parsimony tree obtained from analysis of ITS DNA sequence data.**

The parsimony tree groups our unknown sample (written in blue) with the genus *Escovopsis* (highlighted in green) and is most closely related to *Escovopsis microspore* CBS 135751*.* 

**Figure 2: Optimal maximum-likelihood obtained from analysis of ITS DNA sequence data.**

The maximum-likelihood tree groups our unknown sample (written in blue) with the genus *Escovopsis* (highlighted in green) and is most closely related to *Escovopsis microspore* CBS 135751. The bootstrap values, after 200 iterations, are displayed at each internal node.

**Figure 3:** **Optimal Bayesian obtained from analysis of ITS DNA sequence data.**

The bayesian tree groups our unknown sample (written in blue) with the genus *Escovopsis* (highlighted in green) and is most closely related to *Escovopsis microspore* CBS 135751. The bootstrap values, after 1 million generations, are displayed at each internal node.