**Computer Lab 5 – Phylogenetics I:**

**Maximum Parsimony**

**Conservation Genetics (BIOL 4174 / 5174)**

**Part I – Parsimony**

In this section, you will explore the basic elements of tree building via parsimony. This part of the lab will use Mesquite to build a parsimony tree without the aid of a tree-building algorithm. **Download the Lab 5 files from Blackboard or GitHub and extract them to your Desktop.**

Open the heloderma.nex file in Mesquite (File 🡪 Open File…). The sequences in this file have already been aligned for you.

You will now generate a tree that you will manipulate manually (Taxa&Trees 🡪 New Tree Window 🡪 With Tree to Edit By Hand). You should now see a tree in your window. The order of the taxa in this tree was determined by the order of the taxa in the file. The current branching pattern within the tree is completely arbitrary, thus the tree topology is currently meaningless. Set the program to calculate a simple statistic for the tree: **treelength**

1. Analysis:Tree 🡪 Values for Current Tree
2. Select “Treelength” from the dialog box and click “ok”
3. You should now see a box that reports the treelength of your tree.

* **Do homework question 1**

The basic idea behind parsimony is that the simplest solution is the best solution. The treelength statistic provided by Mesquite is a measure of the number of character state changes that occur within the tree, so you will want to minimize this number. In our case, this is the number of changes from one nucleotide to another (this is why an accurate sequence alignment is so important), although parsimony can just as easily handle morphological characters as well.

Spend no more than 10 minutes moving around branches in your tree using the C:\Program Files (x86)\Mesquite_Folder\images\arrow.giftool. To move branches, hover over a branch with the cursor then click and drag it to connect with the desired branch. The treelength should change with most moves, although you may find some moves to be equally parsimonious. Try to reduce the treelength as much as possible. If one of your moves increases the treelength, you can undo it from the menu bar (Edit 🡪 Undo). **Note**: *undo will only undo your most recent move*.

* **Do homework question 2**

Now, root your tree on the branch leading to the Tuatara (*Sphenodon punctatus*). To do this, use the C:\Program Files (x86)\Mesquite_Folder\mesquite\trees\BasicTreeWindowMaker\reroot.gif tool. Click on this tool, and then click the branch leading to *Sphenodon*. This taxon should now be sister to all the other taxa. Rooting a tree gives the tree **directionality** – it allows us to determine which character states are ancestral and which are derived.

Mesquite can also be used to view ancestral character states. To do this, go to Analysis:Tree 🡪 Trace Character History. In the next window select “Parsimony Ancestral States” and click “ok.” The “Trace Character” box should now appear under your tree. Use the arrows in this box to scroll through the characters. The program is showing you how each individual character (nucleotide position) changes over the tree. Note that some character states are non-variable in your tree.

**Find one example of each of the following:**

* **Autapomorphy**
* **Synapomorphy**
* **Homoplasy – reversal**
* **Homoplasy – convergence**

**Follow the directions for question 3 in the homework document.**

Part II – Maximum Parsimony in MEGA

For the remainder of today’s lab, we will be utilizing MEGA to explore various treebuilding algorithms and assess statistical support for trees. This program is available from <http://www.megasoftware.net/> but has already been installed on your lab computers.

For our first tree, we will use the “Branch & Bound” method. This is guaranteed to find the most parsimonious tree even though it does not search all trees. Rather than exploring and calculating the length of all possible trees as would be done in an exhaustive search, the branch and bound algorithm first eliminates large groups of trees that are suboptimal. This allows for the most parsimonious tree to be found for numbers of taxa for which an exhaustive search cannot be conducted.

Open the file heloderma.meg in MEGA (File 🡪 Open A File/Session…). In the resulting dialog box, make sure “Nucleotide Sequences” is highlighted and click “OK.” When asked if this is protein-coding nucleotide sequence data, select “Yes.” In the next dialog box, make sure “Vertebrate Mitochondrial” is selected, then click “OK.”

Click on the “Phylogeny” button at the top of the screen and choose “Construct/Test Maximum Parsimony Tree(s)”. If asked “Would you like to use the currently active data (heloderma.meg)?” then check the box next to “Remember to reuse currently active data.” And click “Yes.” Then select the following options:

Test of Phylogeny: **None**

Substitutions Type: **Nucleotide**

Gaps/Missing Data Treatment: **Complete Deletion**

Select Codon Positions: **all boxes should be checked**

MP Search Method: **Max-Mini Branch-&-Bound**

When everything has been set, click “Compute” to make a tree. Then, re-root the tree on *Sphenodon* using the Root tool (**Subtree -> Root**) .

* **Do homework questions 4 – 7**

Consensus Trees

You will now calculate a **50% majority rule consensus tree** from the multiple equally parsimonious trees that were found by MEGA. In the window that displays your tree, go to Compute 🡪 Consensus. On the “Cutoff” tab, make sure the “Cut-off Value for Consensus Tree” is set to 50%, then click OK.

* **Do homework questions 8 – 10**

Now, use the same process to make a **strict consensus tree**. The only difference is you will increase the cut-off value from 50% to 100%.

* **Do homework questions 11**

Part III – Bootstrapping

One method of assessing support for your tree is to conduct **nonparametric bootstrapping**. This is the random resampling of your data with replacement to create pseudoreplicate datasets. These datasets are equal in size to your original character matrix. I.e., for each pseudoreplicate the same number of nucleotides is sampled, although some will be sampled multiple times while others are not sampled at all. For each pseudoreplicate, a tree is generated. Once all bootstraps have been conducted, a bootstrap consensus tree is produced. Generally, any node with bootstrap support of less than 50% is collapsed. “High” bootstrap support is subjective, but most generally put the cutoff for a well-supported node at somewhere between 70-80% bootstrap support.

Click on the “Phylogeny” button at the top of the main MEGA window. Once again choose “Construct/Test Maximum Parsimony Tree(s)”. Select the following options:

Test of Phylogeny: **Bootstrap Method**

No. of Bootstrap Replications: **100**

Substitutions Type: **Nucleotide**

Gaps/Missing Data Treatment: **Complete Deletion**

Select Codon Positions: **all boxes should be checked**

MP Search Method: **Max-Mini Branch-&-Bound**

For a tree worthy of publication you would use more bootstrap replicates (typically around 1000) but we will do 100 in the interest of time. Once you have verified all settings, click “Compute.” It will take a few moments for the procedure to finish.

The resulting window will provide you with two tabs. Look at the tab labeled “Original tree,” which contains all of the original equally parsimonious trees found by MEGA, except now nodes are labeled with their bootstrap support values. For Tree#1 do the following:

* **Do homework questions 12 – 15**

Part IV – Heuristic Search Methods

Sometimes a dataset will contain so many taxa that neither an exhaustive search nor the branch and bound method can be used. This necessitates the use of parsimony methods that are incapable of searching all possible trees, and are not guaranteed to find the most parsimonious tree.

Heuristic methods work by first quickly generating a tree using a fast tree-building method (often this is accomplished by building a random tree). Branches are then swapped around on this tree and the tree is rescored after each swap. Trees worse than the starting tree are discarded, while trees that are better are further explored. This continues until better trees are no longer being produced. MEGA offers two commonly used heuristic options – “Tree-Bisection-Reconnection (TBR)” and “Subtree-Pruning-Regrafting (SPR).” Both of these heuristic searches do a good job of finding the most parsimonious trees with our dataset. First, try running this method by constructing a parsimony tree with the following settings:

Test of Phylogeny: **None**

Substitutions Type: **Nucleotide**

Gaps/Missing Data Treatment: **Complete Deletion**

Select Codon Positions: **all boxes should be checked**

MP Search Method: **Tree-Bisection-Reconnection (TBR)**

No. of Initial Trees (random addition): **100**

MP Search level: **1**

Max No. of Trees to Retain: **100**

* **Do homework questions 16 and 17**

Now, repeat this method but decrease the No. of Initial Trees from 100 to 1. Run this method five times to see if you find a tree with a greater treelength (remember: heuristic methods are not always guaranteed to find the shortest tree).

* **Do homework question 18**