

LAB 3: MAXIMUM LIKELIHOOD ESTIMATION USING GARLI

1 Preliminaries

To begin, use a secure shell (ssh, putty, etc.) to connect to one of the following machines:

```
mordor.stat.osu.edu
rohan.stat.osu.edu
gondor.stat.osu.edu
rivendell.stat.osu.edu
shire.stat.osu.edu
```

and login with your username and password. You will need to copy one file for today's exercises. So you should do the following:

```
cp /home/lkubatko/WWW/stat882/garli.conf garli.conf
```

2 About GARLI

GARLI is a program written by Derrick Zwickl that carries out maximum likelihood (ML) inference under several models of sequence substitution. The algorithm used for searching for the ML tree is a genetic algorithm. The program is freely available for download for all platforms. A GUI version is available for PC and Mac systems; we'll use the command line interface available in UNIX, but I will demo the GUI in class briefly. You can find more information about GARLI and download the program onto your own computer here: https://www.nescent.org/wg_garli/Main_Page

3 Running GARLI

To run GARLI in UNIX, you first need to edit the file `garli.conf` – this is the configuration file that contains all the settings for GARLI. At a minimum, you need to change the name of the input file (`datafname`) and the prefix name to use for the output files that GARLI writes (`ofprefix`). To begin, change the input file name to `primates.nex` and change the output prefix to `primates.out`.

We are now ready to run GARLI. To do this, issue the following command at the UNIX prompt:

```
Garli-1.0
```

GARLI should begin running – you will see output being written to the screen.

4 On Your Own – HPV Data

This data set has 38 taxa and consists of 1,518 sites in a single gene. Morrison (Syst. Biol., 2007) examined this data set with GARLI (and several other programs). We can use his results to examine the performance of GARLI. Under the GTR+I+G model, Morrison found that the ML tree has log likelihood -33,790.99. He found that there are two “islands” in the likelihood surface – these are two distinct sets of tree topologies with likelihood scores that are locally optimal but which cannot be transformed into one another through a series of move strategies of a given type. For example, when an analysis method using NNI moves is used, is it often possible to find two different trees such that for each tree every possible NNI move on that tree has a lower likelihood score. Often only one of these trees is globally optimal, but both are local optima and a deterministic search strategy using NNI moves will only find one of the two trees on a single run.

The HPV data set is fun because Morrison found two “GARLI islands” , local optima in the context of allowable moves within GARLI. We will attempt to reproduce those GARLI islands here, and look at how different the trees might be. Open GARLI, and then open the hpv.nex data set. Begin a run with the default settings, except you should change the random number seed (`randseed`) to 87943. Keep the number of searches (`searchreps`) set to 2.

Some things to think about:

- How different are the two trees that were found (remember that GARLI estimates unrooted trees when comparing them)? Since we haven’t used a program to view trees yet, I’ve attached them to this handout.
- What does this mean in terms of the problem itself, in terms of how to carry out analyses, and in terms of how to present results?
- This example also highlights the importance of looking at likelihood scores for multiple trees (in the same way that we might be concerned with the parsimony scores of trees other than the MP tree).