EEOB563 – Assignment #6

Dennis Psaroudakis

April 4th 2019

2

It seems very certain that A and B form a clade (credibility value 100). Assuming this tree is the correct one, A contracted HIV from L1 and then passed it on to B.

```
Clade credibility values:

/------------------------------------------------------------------------ HIVRF (1)
|
|------------------------------------------------------------------------ L2 (3)
|
|------------------------------------------------------------------------ L4 (5)
|
|                            /------------------------------------------ L1 (2)
|                            |
|                            |                   /--------------------- A1 (6)
+                            |                   |
|                   /----64---+                   |         /---------- B1 (7)
|                   |        |         /----99---+----96---+
|                   |        |         |         |         \---------- B2 (8)
|                   |        |         |         |
|         /----97---+        \---100--+         \--------------------- A3 (10)
|         |         |                  |
|         |         |                  \------------------------------ A2 (9)
\----80---+         |
          |         \------------------------------------------------- L5 (11)
          |
          \------------------------------------------------------------ L3 (4)
```

3

There were two models with the same probability of 0.099. See here for more details on the models: https://github.com/Thyra/EEOB563/tree/master/assignments/6/mb/

Model	Posterior Probability	Standard Deviation	Min. Probability	Max. Probability
gtrsubmodel[112212]	0.099	0.008	0.093	0.105

```
5    gtrsubmodel[112312]          0.099          0.015          0.088          0.109
6    gtrsubmodel[123323]          0.081          0.000          0.081          0.081
7    gtrsubmodel[112313]          0.075          0.013          0.065          0.084
8    ------------------------------------------------------------------------------
```

4

The PSRF+ values are all very close to 1. I used `mcmcp ngen=100000 printfreq=100 samplefreq=100 nchains=4` but it was so fast that I could probably have done more generations. I would say the conclusion is independent of the starting tree because the clade has 100% credibility, but I'm not completely sure, so this statement only has about 45% credibility.

```
1                                                95% HPD Interval
2                                               -------------------
3    Parameter      Mean       Variance     Lower       Upper       Median     PSRF+
       Nruns
4    ------------------------------------------------------------------------------------------

5    length[1]     0.133600     0.000471    0.090225    0.174682    0.132569    1.005
           2
6    length[2]     0.003615     0.000006    0.000028    0.008128    0.003151    1.000
           2
7    length[3]     0.001687     0.000003    0.000000    0.004720    0.001223    0.999
           2
8    length[4]     0.019735     0.000037    0.009787    0.032235    0.019310    1.000
           2
9    length[5]     0.018359     0.000036    0.007804    0.030208    0.017806    1.003
           2
10   length[6]     0.004177     0.000006    0.000422    0.008733    0.003701    1.003
           2
11   length[7]     0.002930     0.000005    0.000079    0.006834    0.002416    1.000
           2
12   length[8]     0.004302     0.000006    0.000404    0.009115    0.003800    1.000
           2
13   length[9]     0.004934     0.000008    0.000421    0.010383    0.004374    1.000
           2
14   length[10]    0.007185     0.000010    0.002101    0.013977    0.006690    0.999
           2
```

```
15    length[11]    0.015553    0.000024    0.006808    0.025650    0.015038    0.999
          2
16    length[12]    0.011753    0.000018    0.004477    0.020130    0.011254    1.000
          2
17    length[13]    0.005158    0.000009    0.000735    0.010871    0.004596    1.000
          2
18    length[14]    0.008233    0.000016    0.001471    0.015983    0.007575    0.999
          2
19    length[15]    0.002804    0.000005    0.000014    0.007103    0.002313    1.001
          2
20    length[16]    0.006219    0.000013    0.000382    0.013299    0.005612    1.001
          2
21    length[17]    0.003315    0.000006    0.000005    0.008337    0.002805    1.003
          2
22    length[18]    0.004411    0.000013    0.000004    0.011162    0.003482    0.999
          2
23    length[19]    0.004070    0.000010    0.000006    0.010670    0.003225    1.000
          2
24    length[20]    0.001559    0.000002    0.000003    0.004798    0.001065    0.999
          2
25    length[21]    0.001448    0.000002    0.000006    0.004005    0.001113    1.000
          2
26    length[22]    0.001382    0.000002    0.000003    0.004302    0.000956    1.000
          2
27    length[23]    0.001827    0.000003    0.000005    0.005579    0.001297    0.997
          2
28    length[24]    0.001712    0.000003    0.000006    0.005181    0.001135    1.004
          2
29    length[25]    0.004996    0.000015    0.000001    0.013337    0.004300    0.996
          2
30    length[26]    0.001717    0.000003    0.000006    0.004645    0.001203    1.000
          2
31    ----------------------------------------------------------------------------------

32    + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
33      and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
34      deviation of parameter values within all runs is 0 or when a parameter
35      value (a branch length, for instance) is not sampled in all runs.
```

Here is another convergence statistic that has reached its optimum:

```
Summary statistics for informative taxon bipartitions
   (saved to file "6.2_main.tstat"):

ID   #obs    Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
------------------------------------------------------------------
12  1502    1.000000    0.000000    1.000000    1.000000    2
13  1484    0.988016    0.003766    0.985353    0.990679    2
14  1453    0.967377    0.000942    0.966711    0.968043    2
15  1445    0.962051    0.006591    0.957390    0.966711    2
16  1208    0.804261    0.011299    0.796272    0.812250    2
17   962    0.640479    0.007532    0.635153    0.645806    2
18   720    0.479361    0.007532    0.474035    0.484687    2
19   609    0.405459    0.006591    0.400799    0.410120    2
20   513    0.341545    0.019773    0.327563    0.355526    2
21   491    0.326897    0.014123    0.316911    0.336884    2
22   471    0.313582    0.000942    0.312916    0.314248    2
23   290    0.193076    0.001883    0.191744    0.194407    2
24   239    0.159121    0.008474    0.153129    0.165113    2
25   228    0.151798    0.015065    0.141145    0.162450    2
26   154    0.102530    0.011299    0.094541    0.110519    2
------------------------------------------------------------------
+ Convergence diagnostic (standard deviation of split frequencies)
  should approach 0.0 as runs converge.
```