Phylogenetic Analysis of Molecular Data (Botany 563)

Computer Lab 08: Relaxed Molecular Clock

Learning objectives:

Familiarize with Relaxed Clock analysis in a ML framework using PAUP*, SeqGen (through the SG Runner interface), and r8s.

Learn how to assess confidence intervals for nodal ages in a ML framework.

Using a likelihood ratio test, in the previous lab you rejected the molecular clock hypothesis for the mtDNA *ssu* dataset used by Hibbett 2001 (fourth dataset). Here we will use relaxed the clock to assess nodal ages with confidence intervals to investigate the Old World/NewWorld and Australasian/NewZeland disjuncts discussed by David Hibbett in his paper. We will use age estimate for nodes A and B as calibration points to estimate nodal ages C and D.

Background info:

Fossil calibrations used by Hibbett (2001)		
node	Calibration (Mya)	source
A	200	Prev. study
B	90	Fossil *Archaeomarasmius*
C	100	Gondwana fragmentation
D	80	Australasia-NZ disjunct

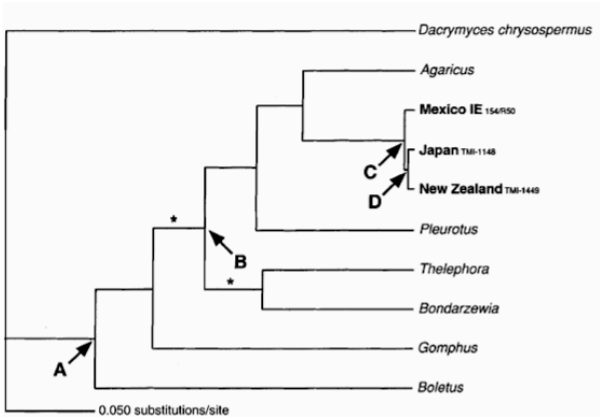

Figure 1. ML tree from Hibbett 2001

Approach:

1. obtain bootrstrapped datasets (using parametric bootstrapping)
2. calculate branchlengths for our ML tree from bootstrapped datasets
3. make trees ultrametric using Penalized Likelihood
4. calculating relative nodal ages for nodes of interest with confidence intervals

DRAG TO YOUR LOCAL DESKTOP the following (from Classroom Scratch):

- hibbett_mt_ssu.nex
- hibbett_ml.tre
- hibbett_mltree_phylip.tre
- hibbett.pdf
- r8s1.71 folder
- lab08_commands.rtf

Task 1. Generate replicate datasets using parametric bootstrapping (Seq-Gen with SG-Runner interface)

a) Launch SG-Runner. Select "new file".
b) Enter the following settings in SG runner window:
    1. Tree file: select "**hibbett_ml_phylip.tre**" (PHYLIP format)
    2. Sequence length (bp): t412
    3. Number of replicates (put 10)
    4. Model: HKY
    5. Transition:transversion ratio: 2
    6. Base freqs = unequal freqs (A=0.35; C=0.156; G=0.208)
    7. Rate Het = GAMMA (alpha = 0.635; ncat=4;PINVAR=0)
    8. Text: check the "insert text block" box, and paste the following PAUP block (you can find the block in the lab08_commands.rtf file):

```
Begin PAUP;
set criterion=likelihood;
lset basefreq=(0.35072560 0.15631319 0.20801963) nst=2 rates=gamma
shape=0.654249 pinvar=0.0;
outgroup Dacrymyces;
gettrees file=hibbett_mltree.tre unrooted=yes warntree=no;
savetrees format=altnexus  file=hibbett_boot_trees append=yes
root=yes brlens=yes maxdecimals=4;
End;
```

c) Hit "Run"
d) The Seq-Gen output will be in a file named "untilted.txt" which will be the input file in the next task. Open this file, look at it, and rename it like "boot_sets".

Task 2. Generating sets of branchlengths for each of the bootrstrap datasets from Seq-Gen.

a) Open the file "boot_sets" in PAUP.
b) Execute the file; this will generate the file "hibbett_boot_trees"
c) Open the file "hibbett_boot_trees"
d) Replace all branchlengths of 0 with 0.0001. This can be done using the "find and replace" function in PAUP, Textedit or any other text editor. You need to do two replacements:
    a. Replace "**:0)**" with "**:0.0001**)"
    b. Replace "**:0,**" with "**:0.0001**,"
e) Save the file and close it.

Task 3. Running a Cross-Validation analysis in r8s.

a) Rename the file "hibbett_boot_trees" as "r8s_cv" and move it to the folder `r8s1.71>bin`
b) Open the file "r8s_cv" and paste the following r8s block at the end of it (this block is also in the "lab08_commands file", under the heading "**r8s command block for Cross Validation**"):

```
Begin r8s;
blformat lengths=persite nsites=412 ultrametric=no;
prune taxon=Albatrellu;
collapse;
mrca nodeA Boletus_sa Lentinu_ed; [root homobasidiomycetes]
mrca nodeB Agaricus_b Bondarzewi; [euagarics+relatives]
mrca nodeC Lentinu_NZ Lentinu_bo; [OW-NW split]
mrca nodeD Lentinu_NZ Lentinu_ed; [Australasia-NZ split]
set checkgradient=no rates=gamma verbose=0;
fixage taxon=nodeA age=200;
fixage taxon=nodeB age=90;
divtime method=pl algorithm=tn crossv=yes cvStart=0 cvInc=0.25 cvNum=15;
cleartrees;
End;
```

c) Open the Terminal (Applications>Utilities>Terminal.app). Something like this should appear:

```
Last login: Wed Mar 25 10:04:52 on console
Welcome to Darwin!
localdisk:~ user$
```

d) Change the directory. **TYPE**: `cd Desktop/r8s1.71/bin`
e) To see what is in that folder type: `ls.` Your file "r8s_cv" should be there.
f) To run r8s type the following in the Terminal: `./r8s -b -f r8s_cv > r8s_cv_out`
WAIT until the prompt "`localdisk:~ user$`" shows up again.
g) Look at the table at the end of the output file "`r8s_cv_out`".
h) **Which value of the examined smoothing parameters (log10 smoothing) would be judged best (lower chi square error that passed the test)? _____**
i) **What does this imply about the rate of evolution across the tree compared to a value of 1.0?**

Task 4. Run a Penalized Likelihood analysis to estimate nodal ages.

a) Rename the file "r8s_cv" as "r8s_pl"
b) In the "r8s_pl" file, substitute the rates block to the following (this block is also in the "lab08_commands file", under the heading "**r8s command block for Penalized Likelihood**"):

```
Begin r8s;
blformat lengths=persite nsites=412 ultrametric=no;
prune taxon=Albatrellu;
collapse;
mrca nodeA Boletus_sa Lentinu_ed; [root homobasidiomycetes]
mrca nodeB Agaricus_b Bondarzewi; [euagarics+relatives]
mrca nodeC Lentinu_NZ Lentinu_bo; [OW-NW split]
mrca nodeD Lentinu_NZ Lentinu_ed; [Australasia-NZ split]
fixage taxon=nodeA age=200;
fixage taxon=nodeB age=90;
set verbose=0 checkgradient=yes penalty=add smoothing=3.5;
divtime method=pl algorithm=tn;
showage shownamed=yes;
describe plot=chronogram;
describe plot=tree_description;
profile taxon=nodeC parameter=age;
profile taxon=nodeD parameter=age;
cleartrees;
End;
```

c) Run the PL analysis. Type: `./r8s -b -f r8s_pl > r8s_pl_out`
d) Look at the output, and complete the following:
   a. Node C: mean=_____; standard deviation: _____.
   b. Node D: mean=_____; standard deviation: _____.

Task 5. Answer the following:

a) a) How do these values compare to Hibbetts'?

b) Do these findings support Hibbett's view that the Old World/New World disjundtion is the result of vicariance?

c) What can you say about the age of *Lentinula* based on your analysis?

Table 1 Molecular clock age estimates of nodes in phylogenetic trees in Fig. 4 (based on nuc-lsu rDNA) and Fig. 5 (based on mt-ssu rDNA). Column 2 gives the node ages in maximum likelihood units (× 100) ± two SE. Columns 3–6 give absolute node age estimates in millions of years ± two SE. Columns 3–6 each represent a set of node ages obtained by fixing the age of one node (in bold) as a calibration point

	Node	ML units	Absolute age estimates			
nuc-lsu rDNA						
	A	127.69 ± 14.94	200	107 ± 25	596 ± 139	1924 ± 450
	B	107.88 ± 11.80	169 ± 37	90	503 ± 110	1625 ± 178
	C	21.44 ± 4.00	34 ± 13	18 ± 7	100	323 ± 121
	D	5.31 ± 1.83	8 ± 6	4 ± 3	25 ± 17	80
mt-ssu rDNA						
	A	131.56 ± 19.21	200	137 ± 20	3132 ± 915	4017 ± 1,173
	B	86.23 ± 12.39	131 ± 38	90	1642 ± 590	2633 ± 757
	C	4.20 ± 1.79	6 ± 5	4 ± 2	100	128 ± 109
	D	2.62 ± 1.43	4 ± 4	3 ± 3	62 ± 68	80