

Phylogenetics Assignment

Part I. Drawing phylogenetic trees (using paper and pencil).

1) Draw the 15 unrooted binary topological trees that could describe the relationship between 5 taxa a, b, c, d, e. Group them naturally according to the relationship they display for the first 4 taxa a, b, c, d.

2) An unrooted tree has vertices v_1, v_2, \dots, v_9 with edges $\{v_1, v_2\}, \{v_1, v_6\}, \{v_1, v_9\}, \{v_2, v_7\}, \{v_2, v_8\}, \{v_3, v_9\}, \{v_4, v_9\}, \{v_5, v_9\}$.

- Without drawing the tree, determine the degree of each vertex.
- Use your answer to part (a) to determine the leaves of the tree.
- Use your answer to part (a) to determine whether the tree is binary.
- Draw the tree to check your work.
- Write this tree in the Newick format.

Part 2. Drawing phylogenetic trees (using computer)

3) Below is a simple code in perl that allows you to draw a phylogenetic tree using postscript commands (<http://www.physics.emory.edu/faculty/weeks//graphics/howtops1.html>). Annotate this code and rewrite it in python. Draw the tree from 2e above. Does it produce the same tree you drew by hand? Re-write the code to draw a tree with the root at the bottom.

```
#!/usr/bin/perl
#drawtree.pl -- draws binary trees (root at top)
#usage : echo '(A(BC)D)(EF) ' drawtree.pl > output.ps
#from Arthur Lesk 2014

print <<EOF;
%!PS-Adobe-
%%BoundingBox : atend
/n /newpath load def
/m /moveto load def
/l /lineto load def
/rm /rmoveto load def
/rl /rlineto load def
/s /stroke load def
1.0 setlinewidth 50 100 translate 2 2 scale
/Helvetica findfont 10 scalefont setfont
EOF

$tree = <>; chop($tree); $_ = reverse($tree); s/[()]/g;
$x = 0; $y = 0;
while ($nd = chop()) {
    print "$x $y m ($nd) stringwidth pop -0.5 mul 0 rm ($nd) show\n";
    $xx{$nd} = $x; $x+=20 ; $yy{$nd} = 10;
}
while ($tree =~ s/\(?([A-Z])([A-Z])\)?/$1/){
    print "n $xx{$1} $yy{$1} m\n";
    ($yy{$1} > $yy{$2}) || {$yy{$1} = $yy{$2}}; $yy{$1} += 20;
    print "$xx{$1} $yy{$1} l $xx{$2} $yy{$1} l $xx{$2} $yy{$2} l s\n";
    $xx{$1} = 0.5*($xx{$1} + $xx{$2});
}
print "n $xx{$tree} $yy{$tree} m 0 20 rl s showpage\n";
$rx = 2*$x + 30; $yt = 2*$yy{$tree} + 146;
print "%BoundingBox : 40 95 $rx $yt\n" ;
```

Part IV. Distance methods using paper and pencil and/or python☺.

Table 1:

Dog	ATG	ACC	AAC	ATT	CGA	AAA	ACC	CAC	CCA	CTA
Cat	ATG	ACC	AAC	ATT	CGA	AAA	TCA	CAC	CCC	CTT
Mouse	ATG	ACA	AAC	ATA	CGA	AAA	ACA	CAC	CCA	TTA
Pig	ATG	ACC	AAC	ATC	CGA	AAA	TCA	CAC	CCA	CTA
Human	ATG	ACC	CCA	ATA	CGC	AAA	ACT	AAC	CCC	CTA

4. Table 1 shows the first 30 bp of the mitochondrial cytochrome b gene for five mammals.

Find the Jukes-Cantor distance for each pair of species ($D(t) = -\ln(1 - (4/3)p)$), where p is uncorrected (observed) dissimilarity). You can write a script to help you with the calculations.

5. Apply the algorithm we used in class to find the UPGMA tree based on JC distances. **Draw this tree and indicate all branch lengths. Write a python program for the UPGMA clustering method.**

Part V. Likelihood methods using paper and pencil ☺.

6. Calculate the likelihood for the following tree under the JC69 model

Taxon:	Sequences:	Topology:
A	AG	
B	AC	
C	AG	

Remember that to calculate the likelihood of a site, you have to consider all possible reconstructions for internal nodes and that the edge lengths are given in terms of expected number of substitutions v , where $v = 3\alpha t$ (for the JC model). You can write a script to help you with the calculations.

GOOD LUCK!