PHYLIP (PHYlogeny Inference Programs)

biol4230 Friday, March 2, 2018
Bill Pearson wrp@virginia.edu 4-2818 Pinn 6-057

PHYLIP via EMBOSS

- Tree building:
  - distance: (f)fitch,(f)kitsch, needs (f)dnadist or (f)protdist first
  - parsimony: (f)dnapars, (f)protpars
  - Likelihood: (f)dnaml, (f)dnamlk, (f)protml
- Tree drawing:
  - (f)drawtree – unrooted
  - (f)drawgram – draws a tree
- Utilities:
  - (f)consense – show consensus tree
  - retree – reroot trees (use interactively)

PHYLIP (PHYlogeny Inference Programs)

- A package of programs developed by Joe Felsenstein; available since 1980
- Written in 'C' for a command line interface
- Available for most popular computers
- Provides a diverse variety of methods for sequence and other data

Phylip 3.69

Advantages

- Free (GNU license)
- Runs on all major platforms
- Good documentation
- Well known/widely used
- Possible to automate
- File formats supported by other packages

Disadvantages

- Much slower than PAUP
- Search strategy less comprehensive
- Primitive command-line interface (user hostile)
- Much file renaming required
- Cannot read NEXUS files

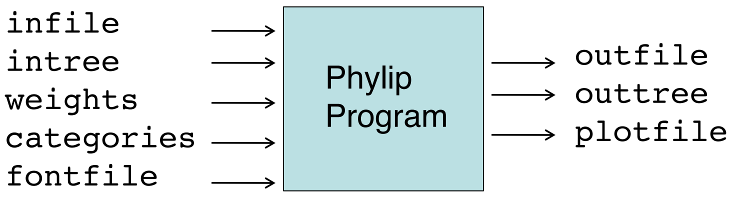
PHYLIP Tree-building programs

- Parsimony:
  - `dnaparse` - parsimony (`protparse`)
  - No branch lengths on trees
- Distance Methods
  - `dnadist, protdist` – produce corrected distance matrices
  - `fitch, kitsch` – Fitch-Margoliash distance (clock, kitsch) trees from distances
  - `Neighbor` – Neighbor-joining trees (no explicit optimization criterion)

PHYLIP Tree-building programs

- Maximum Likelihood
  - `dnaml, dnamlk` - DNA maximum likelihood
  - `proml, promlk` - protein maximum likelihood
  - *mlk methods assume evolutionary clock (all branches end at same level (time)

PHYLIP Program Data/Output



- The phylip programs re-use the same file names: "infile", "outfile", every time a program is used. In current versions, if the input file is not present, it is prompted for, and if the output file is present, one is warned before over-writing it.
- However, it is easy to analyse the wrong data (old "infile") and over write (or mis-name) the output file.
- Develop a protocol for ensuring that file names make sense. NEVER use infile and outfile, outree. This can be difficult. Scripts help.

PHYLIP via EMBOSS

- EMBOSS (European Molecular Biology lab Open Software Suite)
  - command line options
  - interactive when needed (sometimes annoying)
  - use –help
- EMBOSS PHYLIP:
  - f+PHYLIP name: fdnadist, fconsense, ffitch, fkitsch, etc

PHYLIP sequence format (interleaved)

Number of taxa ↘ Length of alignment ↙

```
7 112
Bovine  CCAAACCTGT CCCCACCATC TAACACCAAC CCACATATAC AAGCTAAACC AAAAATACCA
Mouse   CCAAAAAAAC ATCCAAACAC CAACCCCAGC CCTTACGCAA TAGCCATACA AAGAATATTA
Gibbon  CTATACCCAC CCAACTCGAC CTACACCAAT CCCCACATAG CACACAGACC AACAACCTCC
Orang   CCCCACCCGT CTACACCAGC CAACACCAAC CCCCACCTAC TATACCAACC AATAACCTCT
Gorilla CCCCATTTAT CCATAAAAAC CAACACCAAC CCCCATCTAA CACACAAACT AATGACCCCC
Chimp   CCCCATCCAC CCATACAAAC CAACATTACC CTCCATCCAA TATACAAACT AACAACCTCC
Human   CCCCACTCAC CCATACAAAC CAACACCACT CTCCACCTAA TATACAAATT AATAACCTCC

        CCCCAGCCCA ACACCCTTCC ACAAATCCTT AATATACGCA CCATAAATAA CA
        TCCCACCAAA TCACCCTCCA TCAAATCCAC AAATTACACA ACCATTAACC CA
        GCACGCCAAG CTCTCTACCA TCAAACGCAC AACTTACACA TACAGAACCA CA
        ACACCCTAAG CCACCTTCCT CAAAATCCAA AACCCACACA ACCGAAACAA CA
        ACACCTCAAT CCACCTCCCC CCAAATACAC AATTCACACA AACAATACCA CA
        ACATCTTGAC TCGCCTCTCT CCAAACACAC AATTCACGCA AACAACGCCA CA
        ACACCTTAAC TCACCTTCTC CCAAACGCAC AATTCGCACA CACAACGCCA CA
```

use EMBOS `seqret` to convert to PHYLIP format (`-osformat2 phylip`)

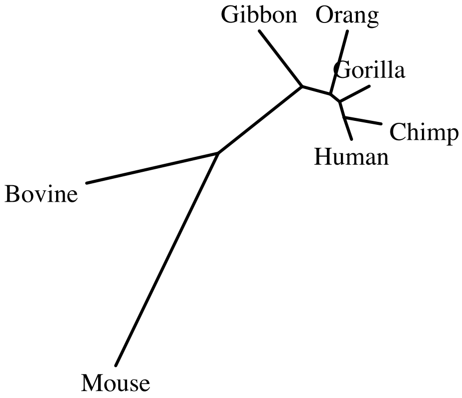
PHYLIP Tree representation (NEWICK)

Taxa label

Branch Length

```
(Mouse:0.87231,Bovine:0.49807,(Gibbon:0.25930,(Orang:0.24
166, (Gorilla:0.12322,(Chimp:0.13846,
Human:0.08571):0.06026):0.04405):0.10815):0.39538);
(Mouse:0.87558,Bovine:0.49718,(Gibbon:0.25698,(Orang:0.24
477, ((Gorilla:0.16328,Chimp:0.13802):0.01842,
Human:0.08495):0.06610):0.10637):0.39287);
(Mouse:0.87819,Bovine:0.49461,(Gibbon:0.25837,(Orang:0.24
161, (Chimp:0.13941,(Gorilla:0.16639,
Human:0.09533):0.00616):0.06709):10938):0.39630);
```

PHYLIP Tree representation (NEWICK)



```
(Mouse:0.87231,Bovine:0.49807,(Gibbon:0.25930,(Orang:0.24
166, (Gorilla:0.12322,(Chimp:0.13846,
Human:0.08571):0.06026):0.04405):0.10815):0.39538);
```

Tree-analysis/display

- Tree comparison:
  - `(f)consense` – Calculate consensus tree from bootstraps
  - `(f)treedist` – compare trees by "partition distance"
- Manipulation
  - `retree` – flip nodes, re-root, re-arrange – run interactively
- Display
  - `(f)drawgram` – draw "tree-like" tree
  - `(f)drawtree` – draw unrooted tree

Running PHYLIP

`infile`
`gstm_n.phy`

```
 15 675
GTM1_HUMAN    ---------- --ATGCCCAT GATACTGGGG TACTGGGACA TCCGCGGGCT
GTM2_HUMAN    ---------- --ATGCCCAT GACACTGGGG TACTGGAACA TCCGCGGGCT
GTM3_HUMAN    ATGTCGTGCG AGTCGTCTAT GGTTCTCGGG TACTGGGATA TTCGTGGGCT
GTM4_HUMAN    ---------- --ATGTCCAT GACACTGGGG TACTGGGACA TCCGCGGGCT
GTM5_HUMAN    ---------- --ATGCCCAT GACTCTGGGG TACTGGGACA TCCGTGGGCT
GTM1_MOUSE    ---------- --ATGCCTAT GATACTGGGA TACTGGAACG TCCGCGGACT
GTM2_MOUSE    ---------- --ATGCCTAT GACACTAGGT TACTGGGACA TCCGTGGGCT
GTM3_MOUSE    ---------- --ATGCCTAT GACACTGGGC TATTGGAACA CCCGCGGACT
GTM5_MOUSE    ATGTCATCCA AGTCT---AT GGTTCTGGGT TACTGGGATA TCCGCGGGCT
GTM1_RAT      ---------- --ATGCCTAT GATACTGGGA TACTGGAACG TCCGCGGGCT
GTM2_RAT      ---------- --ATGCCTAT GACACTGGGT TACTGGGACA TCCGTGGGCT
GTM3_RAT      ---------- --ATGCCCAT GACACTGGGT TACTGGGACA TCCGTGGGCT
GTMU_CRILO    ---------- --ATGCCTAT GATACTGGGA TACTGGAATG TCCGCGGTCT
GTMU_MESAU    ---------- --ATGCCTGT GACACTGGGT TACTGGGACA TCCGTGGGCT
GTM2_CHICK    ---------- --ATGGTGGT CACGTTGGGT TATTGGGACA TCCGCGGGTT

              GGCCCACGCC ATCCGCCTGC TCCTGGAATA CACAGACTCA AGCTATGAGG
              GGCCCATTCC ATCCGCCTGC TCCTGGAATA CACAGACTCA AGCTACGAGG
              GGCGCACGCC ATCCGCCTGC TCCTGGAGTT CACGGATACC TCTTATGAGG
              GGCCCACGCC ATCCGCCTGC TCCTGGAATA CACAGACTCA AGCTACGAGG
              GGCCCACGCC ATCCGCTTGC TCCTGGAATA CACAGACTCA AGCTATGTGG
              GACACACCCG ATCCGCATGC TCCTGGAATA CACAGACTCA AGCTATGATG
              GGCTCACGCC ATCCGCCTGC TCCTGGAATA CACAGACACA AGCTATGAGG
              GACTCACTCC ATCCGCTTGC TCCTGGAATA CACAGATTCA AGCTATGAGG
              GGCTCATGCT ATCCGCATGC TTCTGGAGTT TACTGATACC AGCTATGAGG
              GACACACCCG ATCCGCCTGC TCCTGGAATA CACAGACTCA AGCTATGAGG
              GGCTCACGCC ATTCGCCTGT TCCTGGAGTA TACAGACACA AGCTATGAGG
              AGCGCATGCC ATCCGCCTGC TCCTGGAATA CACAGACTCG AGCTATGAGG
              GACAAACCCC ATCCGCCTGC TCCTGGAATA CACAGACTCA AGCTATGAGG
              GGCTCATGCC ATCCGCCTGC TCTTGGAGTA CACAGACACA AGCTATGAGG
              GGCCCACGCC ATCCGCCTGC TGCTGGAGTA CACCGAGACC CCCTACCAGG
```

Running PHYLIP - `dnaml`

```
$ fdnaml -help
   Standard (Mandatory) qualifiers:
  [-sequence]          seqsetall  File containing one or more sequence
                                  alignments
  [-intreefile]        tree       Phylip tree file (optional)
  [-outfile]           outfile    [*.fdnaml] Phylip dnaml program output file

   Additional (Optional) qualifiers (* if not always prompted):
   -ncategories        integer    [1] Number of substitution rate categories
                                  (Integer from 1 to 9)
   -weights            properties Weights file
*  -njumble            integer    [0] Number of times to randomise (Integer 0
                                  or more)
*  -seed               integer    [1] Random number seed between 1 and 32767
                                  (must be odd) (Integer from 1 to 32767)
*  -global             boolean    [N] Global rearrangements
   -outgrno            integer    [0] Species number to use as outgroup (Integer
•   -outtreefile        outfile    [*.fdnaml] Phylip tree output file  (optional)

General qualifiers:
   -help               boolean    Report command line options. More
                                  information on associated and general
                                  qualifiers can be found with -help -verbose
```

Running PHYLIP – `(f)dnaml`

```
Nucleic acid sequence Maximum Likelihood method, version 3.63
Settings for this run:
  U                 Search for best tree?  Yes
  T        Transition/transversion ratio:  2.0000
  F       Use empirical base frequencies?  Yes
  C               One category of sites?  Yes
  R           Rate variation among sites?  constant rate
  W                       Sites weighted?  No
  S        Speedier but rougher analysis?  Yes
  G               Global rearrangements?  No
  J   Randomize input order of sequences?  No. Use input order
  O                        Outgroup root?  Yes, at sequence number 15
  M           Analyze multiple data sets?  No
  I          Input sequences interleaved?  Yes
  0   Terminal type (IBM PC, ANSI, none)?  ANSI
  1    Print out the data at start of run  No
  2  Print indications of progress of run  Yes
  3                        Print out tree  Yes
  4       Write out trees onto tree file?  Yes
  5   Reconstruct hypothetical sequences?  No
  Y to accept these or type the letter for one to change
j
Random number seed (must be odd)?
123
Number of times to jumble?
5
```

Running PHYLIP – `(f)dnaml`

```
Nucleic acid sequence Maximum Likelihood method, version 3.63
Empirical Base Frequencies:
   A       0.25824                                          Ln Likelihood = -4967.04025
   C       0.25662
   G       0.25997
  T(U)     0.22516                                           Betwn   And         Length   Approx. Confid. Limits
Transition/transversion ratio =   2.000000                   -----   ---         ------   ------- ------- ------

                +---GTM3_MOUSE                               11   GTM2_CHICK   0.31594  (  0.25746, 0.37441) **
             +--8                                            11      3         0.08672  (  0.05406, 0.11939) **
             |  |  +--GTMU_CRILO                              3      4         0.02793  (  0.01168, 0.04422) **
             |  +--9                                          4      2         0.02634  (  0.01094, 0.04173) **
             |     |  +-GTM1_MOUSE                            2      8         0.03562  (  0.01808, 0.05315) **
          +--2     +-12                                       8   GTM3_MOUSE   0.08136  (  0.05732, 0.10551) **
          |  |        +-GTM1_RAT                              8      9         0.01699  (  0.00496, 0.02902) **
          |  |                                                9   GTMU_CRILO   0.05525  (  0.03583, 0.07467) **
          |  |  +---GTMU_MESAU                                9     12         0.01836  (  0.00619, 0.03053) **
       +--4  +--5                                            12   GTM1_MOUSE   0.03072  (  0.01641, 0.04505) **
       |  |     |  +--GTM2_RAT                               12   GTM1_RAT     0.03335  (  0.01837, 0.04833) **
       |  |     +--1                                          2      5         0.04458  (  0.02524, 0.06391) **
       |  |        +GTM2_MOUSE                                5   GTMU_MESAU   0.07139  (  0.04896, 0.09383) **
       |  |                                                   5      1         0.02084  (  0.00814, 0.03354) **
  +----3  +----GTM3_RAT                                       1   GTM2_RAT     0.04346  (  0.02629, 0.06055) **
  |    |                                                      1   GTM2_MOUSE   0.01543  (  0.00448, 0.02638) **
  |    |  +---GTM2_HUMAN                                      4   GTM3_RAT     0.08214  (  0.05771, 0.10667) **
  |    |  |                                                   3      6         0.02335  (  0.00714, 0.03966) **
  |    +--6     +--GTM4_HUMAN                                 6   GTM2_HUMAN   0.07147  (  0.04929, 0.09368) **
  |       |  +-10                                             6      7         0.00694  (     zero, 0.01543) *
  |       +--7  +-GTM1_HUMAN                                  7     10         0.01316  (  0.00296, 0.02336) **
  |          |                                               10   GTM4_HUMAN   0.05475  (  0.03560, 0.07398) **
  |          +---GTM5_HUMAN                                  10   GTM1_HUMAN   0.03047  (  0.01614, 0.04490) **
  |                                                           7   GTM5_HUMAN   0.06349  (  0.04281, 0.08419) **
  |       +----GTM5_MOUSE                                    11     13         0.13085  (  0.09317, 0.16853) **
 11------13                                                  13   GTM5_MOUSE   0.07982  (  0.05403, 0.10560) **
  |       +---GTM3_HUMAN                                     13   GTM3_HUMAN   0.06202  (  0.03845, 0.08568) **
  |
  +------------------GTM2_CHICK                               *  = significantly positive, P < 0.05
                                                              ** = significantly positive, P < 0.01
remember: (although rooted by outgroup) this is an unrooted tree!
```

Running PHYLIP – `(f)dnapars`

```
DNA parsimony algorithm, version 3.63

Setting for this run:
  U                 Search for best tree?  Yes
  S                        Search option?  More thorough search
  V              Number of trees to save?  10000
  J   Randomize input order of sequences?  No. Use input order
  O                        Outgroup root?  Yes, at sequence number 15
  T              Use Threshold parsimony?  No, use ordinary parsimony
  N           Use Transversion parsimony?  No, count all steps
  W                       Sites weighted?  No
  M           Analyze multiple data sets?  No
  I          Input sequences interleaved?  Yes
  0   Terminal type (IBM PC, ANSI, none)?  ANSI
  1    Print out the data at start of run  No
  2  Print indications of progress of run  Yes
  3                        Print out tree  Yes
  4          Print out steps in each site  No
  5  Print sequences at all nodes of tree  No
  6       Write out trees onto tree file?  Yes

  Y to accept these or type the letter for one to change
Y
```

Running PHYLIP – `(f)dnapars`

```
DNA parsimony algorithm, version 3.63                requires a total of    913.000

     3 trees in all found                              between      and       length
                                                       -------      ---       ------
  +-----------GTM2_CHICK                                  2      GTM2_CHICK   0.203366
  |                                                       2           8   0.131035
  |       +---GTM5_MOUSE                                  8      GTM5_MOUSE   0.075672
  2-------8                                               8      GTM3_HUMAN   0.061172
  |       +---GTM3_HUMAN                                  2           5   0.085438
  |                                                       5          11   0.026262
  |       +---GTM3_RAT                                   11      GTM3_RAT     0.067351
  |       |                                              11           6   0.027000
  |       |     +--GTMU_MESAU                             6          13   0.038716
  |    +-11  +-13                                        13      GTMU_MESAU   0.062522
  |    |  |  |  |  +-GTM2_RAT                            13          10   0.020370
  |    |  |  |  +-10                                     10      GTM2_RAT     0.037725
  |    |  +--6     +GTM2_MOUSE                           10      GTM2_MOUSE   0.017584
  |    |     |                                            6           7   0.032519
  |    |     |  +---GTM3_MOUSE                            7      GTM3_MOUSE   0.067937
  |    |     +--7                                         7          12   0.020952
  |    |        |  +-GTMU_CRILO                          12      GTMU_CRILO   0.049136
  +----5        +-12                                     12           9   0.018272
       |           |  +GTM1_RAT                           9      GTM1_RAT     0.031111
       |           +--9                                   9      GTM1_MOUSE   0.028148
       |              +GTM1_MOUSE                         5           1   0.030898
       |                                                  1           4   0.009778
       |     +--GTM5_HUMAN                                4      GTM5_HUMAN   0.056824
       |  +--4                                            4      GTM2_HUMAN   0.061695
       |  |  +--GTM2_HUMAN                                1           3   0.013210
       +--1                                               3      GTM4_HUMAN   0.047152
          |  +--GTM4_HUMAN                                3      GTM1_HUMAN   0.030750
          +--3
             +-GTM1_HUMAN
```

`(f)dnapars` – three alternate trees

```
(GTM2_CHICK:0.20337,(GTM5_MOUSE:0.07567,GTM3_HUMAN:0.06117):0.13103,
((GTM3_RAT:0.06735,((GTMU_MESAU:0.06252,(GTM2_RAT:0.03772,GTM2_MOUSE:0.01758):0.02037):0.
03872,
(GTM3_MOUSE:0.06794,(GTMU_CRILO:0.04914,(GTM1_RAT:0.03111,GTM1_MOUSE:0.02815):0.01827):0.
02095):0.03252):0.02700):0.02626,
((GTM5_HUMAN:0.05682,GTM2_HUMAN:0.06169):0.00978,(GTM4_HUMAN:0.04715,
GTM1_HUMAN:0.03075):0.01321):0.03090):0.08544)[0.3333];

(GTM2_CHICK:0.19762,(GTM5_MOUSE:0.07698,GTM3_HUMAN:0.05942):0.13647,
(((GTMU_MESAU:0.06103,(GTM2_RAT:0.03807,GTM2_MOUSE:0.01723):0.02135):0.03741,
(GTM3_MOUSE:0.06916,(GTMU_CRILO:0.04806,(GTM1_RAT:0.03111,GTM1_MOUSE:0.02815):0.01935):0.
02106):0.03236):0.02522,
(GTM3_RAT:0.06150,(GTM2_HUMAN:0.05333,(GTM5_HUMAN:0.05213,(GTM4_HUMAN:0.04975,
GTM1_HUMAN:0.02815):0.01713):0.01605):0.04058):0.02860):0.08532)[0.3333];

(GTM2_CHICK:0.20335,(GTM5_MOUSE:0.07591,GTM3_HUMAN:0.06098):0.13099,
((GTM3_RAT:0.06487,((GTMU_MESAU:0.06237,(GTM2_RAT:0.03787,GTM2_MOUSE:0.01744):0.02037):0.
03904,
(GTM3_MOUSE:0.06806,(GTMU_CRILO:0.04899,(GTM1_RAT:0.03111,GTM1_MOUSE:0.02815):0.01842):0.
02098):0.03254):0.02944):0.02617,
(GTM2_HUMAN:0.05754,(GTM5_HUMAN:0.05427,(GTM4_HUMAN:0.05030,GTM1_HUMAN:0.02760):0.01481):
0.01128):0.03306):0.08668)[0.3333];
```

Running PHYLIP – `distance methods`

- Distance methods do not work on alignments, they work on distances
  - take alignment and build (corrected) distance matrix `fdnadist, fprotdist`
  - take distance matrix, build tree using `ffitch` (no –evolutionary clock), or `fkitsch` (clock-like tree)
  - `fneighbor` for speed

Running PHYLIP – `(f)dnadist`

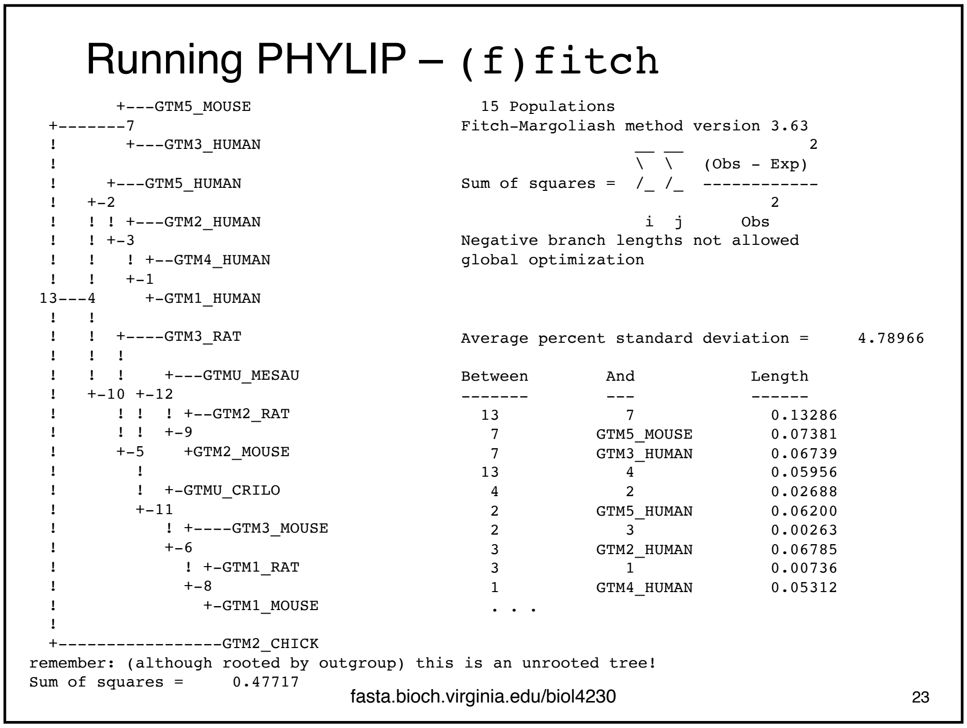
```
Nucleic acid sequence Distance Matrix program, version 3.63

Settings for this run:
  D  Distance (F84, Kimura, Jukes-Cantor, LogDet)?  F84
  G          Gamma distributed rates across sites?  No
  T                 Transition/transversion ratio?  2.0
  C            One category of substitution rates?  Yes
  W                        Use weights for sites?  No
  F                Use empirical base frequencies?  Yes
  L                      Form of distance matrix?  Square
  M                   Analyze multiple data sets?  No
  I                  Input sequences interleaved?  Yes
  0            Terminal type (IBM PC, ANSI, none)?  ANSI
  1             Print out the data at start of run  No
  2           Print indications of progress of run  Yes

  Y to accept these or type the letter for one to change
y
                                                  Distances calculated for species
                                                      GTM1_HUMAN   ..............
                                                      GTM2_HUMAN   .............
                                                      GTM3_HUMAN   ............
                                                      GTM4_HUMAN   ...........
                                                      GTM5_HUMAN   ..........
                                                      GTM1_MOUSE   .........
                                                      GTM2_MOUSE   ........
                                                      GTM3_MOUSE   .......
                                                  . . .
                                                  Distances written to file "gstm_n.ddist"
                                                  Done.
```

Running PHYLIP – `(f)dnadist`

```
 15
GTM1_HUMAN  0.000000  0.111515  0.328043  0.084938  0.098515  0.202847
  0.160670  0.222157  0.323212  0.195992  0.188005  0.176254  0.169073
  0.202499  0.472135
GTM2_HUMAN  0.111515  0.000000  0.370425  0.122881  0.135281  0.234489
  0.198432  0.246131  0.367307  0.220479  0.235718  0.162609  0.200569
  0.245624  0.499002
GTM3_HUMAN  0.328043  0.370425  0.000000  0.330864  0.337744  0.395844
  0.350801  0.407140  0.141206  0.397266  0.389013  0.385259  0.364146
  0.386434  0.489052
GTM4_HUMAN  0.084938  0.122881  0.330864  0.000000  0.131796  0.233678
  0.187505  0.236442  0.337068  0.235722  0.213963  0.182756  0.204816
  0.204302  0.452330
GTM5_HUMAN  0.098515  0.135281  0.337744  0.131796  0.000000  0.230120
  0.186003  0.230817  0.353029  0.215696  0.218532  0.174287  0.201916
  0.216947  0.470660
GTM1_MOUSE  0.202847  0.234489  0.395844  0.233678  0.230120  0.000000
  0.160969  0.116636  0.395293  0.062703  0.200109  0.200296  0.105091
  0.202873  0.486157
GTM2_MOUSE  0.160670  0.198432  0.350801  0.187505  0.186003  0.160969
  0.000000  0.172174  0.370651  0.159042  0.058864  0.178584  0.146716
  0.103994  0.474313
. . .
```

Running PHYLIP – `(f)fitch`

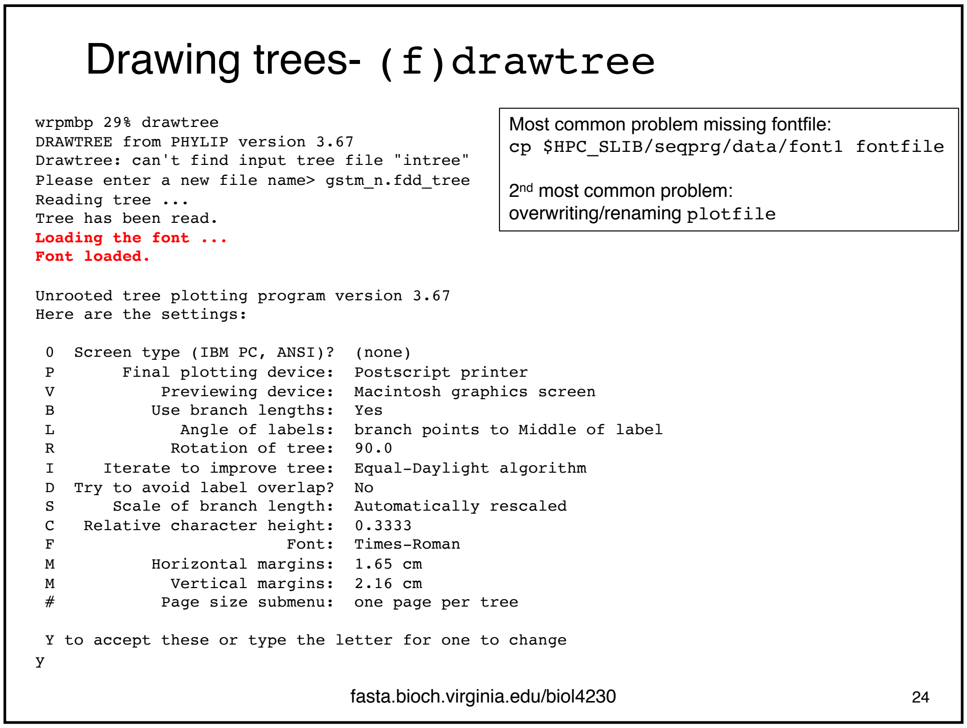
```
Fitch-Margoliash method version 3.63

Settings for this run:
  D      Method (F-M, Minimum Evolution)?  Fitch-Margoliash
  U                 Search for best tree?  Yes
  P                                Power?  2.00000
  -      Negative branch lengths allowed?  No
  O                        Outgroup root?  Yes, at species number 15
  L         Lower-triangular data matrix?  No
  R         Upper-triangular data matrix?  No
  S                        Subreplicates?  No
  G                Global rearrangements?  Yes
  J     Randomize input order of species?  No. Use input order
  M           Analyze multiple data sets?  No
  0   Terminal type (IBM PC, ANSI, none)?  ANSI
  1    Print out the data at start of run  No
  2  Print indications of progress of run  Yes
  3                        Print out tree  Yes
  4       Write out trees onto tree file?  Yes

  Y to accept these or type the letter for one to change
y
```



Running PHYLIP – `(f)fitch`

```
        +---GTM5_MOUSE
 +-------7
 !        +---GTM3_HUMAN
 !
 !      +---GTM5_HUMAN
 !   +-2
 !   ! ! +---GTM2_HUMAN
 !   ! +-3
 !   !   ! +--GTM4_HUMAN
 !   !   +-1
13---4     +-GTM1_HUMAN
 !   !
 !   !  +----GTM3_RAT
 !   !  !
 !   !  !    +---GTMU_MESAU
 !   +-10 +-12
 !      ! !  ! +--GTM2_RAT
 !      ! !  +-9
 !      +-5    +GTM2_MOUSE
 !        !
 !        !  +-GTMU_CRILO
 !        +-11
 !           ! +----GTM3_MOUSE
 !           +-6
 !             ! +-GTM1_RAT
 !             +-8
 !               +-GTM1_MOUSE
 !
 +-----------------GTM2_CHICK
remember: (although rooted by outgroup) this is an unrooted tree!
Sum of squares =     0.47717
```

```
  15 Populations
Fitch-Margoliash method version 3.63
                  __ __                 2
                  \  \   (Obs - Exp)
Sum of squares =  /_ /_  ------------
                                2
                   i  j      Obs
Negative branch lengths not allowed
global optimization



Average percent standard deviation =     4.78966

Between        And            Length
-------        ---            ------
  13             7              0.13286
   7          GTM5_MOUSE        0.07381
   7          GTM3_HUMAN        0.06739
  13             4              0.05956
   4             2              0.02688
   2          GTM5_HUMAN        0.06200
   2             3              0.00263
   3          GTM2_HUMAN        0.06785
   3             1              0.00736
   1          GTM4_HUMAN        0.05312
    . . .
```



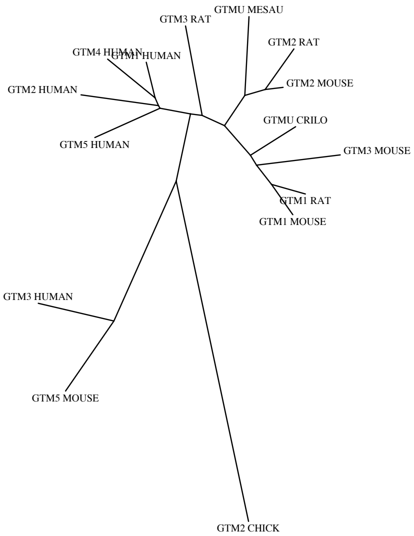
Drawing trees- `(f)drawtree`

Most common problem missing fontfile:
`cp $HPC_SLIB/seqprg/data/font1 fontfile`

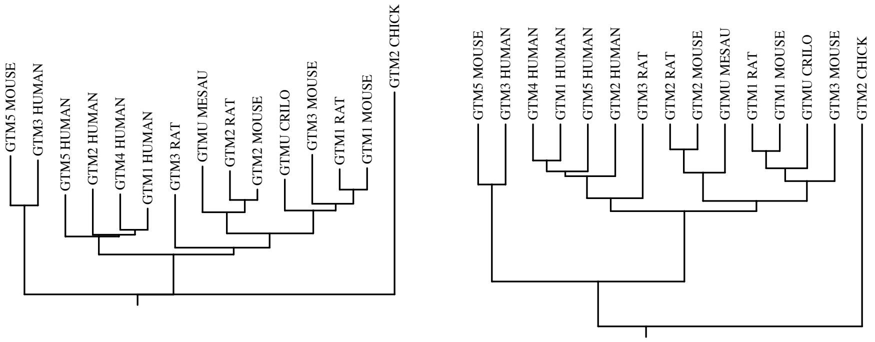
2nd most common problem:
overwriting/renaming `plotfile`

```
wrpmbp 29% drawtree
DRAWTREE from PHYLIP version 3.67
Drawtree: can't find input tree file "intree"
Please enter a new file name> gstm_n.fdd_tree
Reading tree ...
Tree has been read.
Loading the font ...
Font loaded.

Unrooted tree plotting program version 3.67
Here are the settings:

 0  Screen type (IBM PC, ANSI)?  (none)
 P       Final plotting device:  Postscript printer
 V           Previewing device:  Macintosh graphics screen
 B          Use branch lengths:  Yes
 L             Angle of labels:  branch points to Middle of label
 R            Rotation of tree:  90.0
 I     Iterate to improve tree:  Equal-Daylight algorithm
 D  Try to avoid label overlap?  No
 S      Scale of branch length:  Automatically rescaled
 C   Relative character height:  0.3333
 F                        Font:  Times-Roman
 M          Horizontal margins:  1.65 cm
 M            Vertical margins:  2.16 cm
 #           Page size submenu:  one page per tree

 Y to accept these or type the letter for one to change
y
```

Drawing trees- `(f)drawtree`



Drawing trees- `(f)drawgram`



fitch

kitcsh -
(evolutionary clock)

Evaluating trees- `(f)consense`

```
Consensus tree program, version 3.63

Settings for this run:
 C         Consensus type (MRe, strict, MR, Ml):  Majority rule (extended)
 O                                Outgroup root:  Yes, at species number 15
 R                Trees to be treated as Rooted:  No
 T           Terminal type (IBM PC, ANSI, none):  ANSI
 1                Print out the sets of species:  Yes
 2         Print indications of progress of run:  Yes
 3                               Print out tree:  Yes
 4               Write out trees onto tree file:  Yes

Are these settings correct? (type Y or the letter for one to change)
y
```

Evaluating trees- `(f)consense`

```
Consensus tree program, version 3.63

Species in order:

  1. GTM5 MOUSE
  2. GTM3 HUMAN
  3. GTM5 HUMAN
  4. GTM2 HUMAN
  5. GTM4 HUMAN
  6. GTM1 HUMAN
  7. GTM3 RAT
  8. GTMU MESAU
  9. GTM2 RAT
  10. GTM2 MOUSE

  11. GTMU CRILO
  12. GTM3 MOUSE
  13. GTM1 RAT
  14. GTM1 MOUSE
  15. GTM2 CHICK
```

```
Sets included in the consensus tree

Set (species in order)     How many times out of 3.00

.......... ****.             3.00
..****.... .....             3.00
.......... ..**.             3.00
........** .....             3.00
.......*** ****.             3.00
..******** ****.             3.00
....**.... .....             3.00
.......*** .....             3.00
**........ .....             3.00
......**** ****.             2.67
.......... .***.             2.00
...***.... .....             2.00


Sets NOT included in consensus tree:

Set (species in order)     How many times out of 3.00

.......... *.**.             1.00
..*.**.... .....             0.67
..**...... .....             0.33
..*****... .....             0.33
```

Evaluating trees- `(f)consense`

```
Extended majority rule consensus tree

CONSENSUS TREE:
the numbers on the branches indicate the number
of times the partition of the species into the two sets
which are separated by that branch occurred
among the trees, out of   3.00 trees
```

```
                                                 +------GTM1 RAT
                                          +--3.0-|
                                   +--2.0-|      +------GTM1 MOUSE
                                   |      |
                            +--3.0-|      +-------------GTM3 MOUSE
                            |      |
                     +--3.0-|      +--------------------GTMU CRILO
                     |      |
                     |      |             +-------------GTMU MESAU
                     |      +---------3.0-|
              +--2.7-|                    |      +------GTM2 RAT
              |      |                    +--3.0-|
              |      |                           +------GTM2 MOUSE
              |      |
              |      +----------------------------------GTM3 RAT
       +--3.0-|
       |      |                                  +------GTM1 HUMAN
       |      |                           +--3.0-|
       |      |                    +--2.0-|      +------GTM4 HUMAN
       |      |                    |      |
+------|      +----------------3.0-|      +-------------GTM2 HUMAN
|      |                           |
|      |                           +--------------------GTM5 HUMAN
|      |
|      |                                         +------GTM3 HUMAN
|      +-------------------------------------3.0-|
|                                                +------GTM5 MOUSE
|
+-------------------------------------------------------GTM2 CHICK
```

```
remember: (though rerooted by outgroup) this is an unrooted tree!
```

Putting it all together, the User tree

- The problem:
  - the `(f)consense` program produces the best consensus tree, but the branches reflect the consensus frequencies, not the evolutionary branch lengths
- The solution:
  - give consensus tree to `fdnaml` or `ffitch` using the `'U'` user tree option – calculates branches for a single tree, does not do a search (fast)

User tree – (f)`dnaml`

```
Nucleic acid sequence Maximum Likelihood method, version 3.63

Settings for this run:
  U                 Search for best tree?  No, use user trees in input file
  L          Use lengths from user trees?  No
  T        Transition/transversion ratio:  2.0000
  F       Use empirical base frequencies?  Yes
  C                One category of sites?  Yes
  R           Rate variation among sites?  constant rate
  W                       Sites weighted?  No
  O                        Outgroup root?  No, use as outgroup species  1
  M           Analyze multiple data sets?  No
  I          Input sequences interleaved?  Yes
  0   Terminal type (IBM PC, ANSI, none)?  ANSI
  1    Print out the data at start of run  No
  2  Print indications of progress of run  Yes
  3                        Print out tree  Yes
  4       Write out trees onto tree file?  Yes
  5   Reconstruct hypothetical sequences?  No

  Y to accept these or type the letter for one to change
```

`Asks for infile` (alignment) and `intree` (consensus tree)

User tree – `dnaml`

```
User-defined tree:
                      +-GTM1_RAT
                   +--7
                +--6  +-GTM1_MOUSE
                |  |
             +--5  +----GTM3_MOUSE
             |  |
          +--4  +--GTMU_CRILO
          |  |
          |  |  +---GTMU_MESAU
          |  +--8
       +--3     |  +--GTM2_RAT
       |  |     +--9
       |  |        +GTM2_MOUSE
       |  |
       |  +----GTM3_RAT
  +----2
  |    |        +-GTM1_HUMAN
  |    |     +-12
  |    |  +-11  +--GTM4_HUMAN
  |    |  |  |
  |    +-10  +---GTM2_HUMAN
  |       |
  |       +---GTM5_HUMAN
  |
  |       +---GTM3_HUMAN
  1------13
  |       +----GTM5_MOUSE
  |
  +------------------GTM2_CHICK
remember: (although rooted by outgroup) this is an unrooted tree!
```

Consensus tree DNAML:
`Ln Likelihood = -4977.65455`

Original best DNAML:
`Ln Likelihood = -4967.04025`

Phylip for dummies

- Programs for Parsimony, Distance, and Maximum Likelihood
- `infile/outfile/outtree/intree`
  - either always change, or never use
  - Use EMBOS (f) programs
- `(f)consense` to build consensus tree (but invalid branch lengths)
- User tree to calculate branch lengths for consensus tree
- `(f)drawtree` for non-trees, `(f)drawgram` for trees