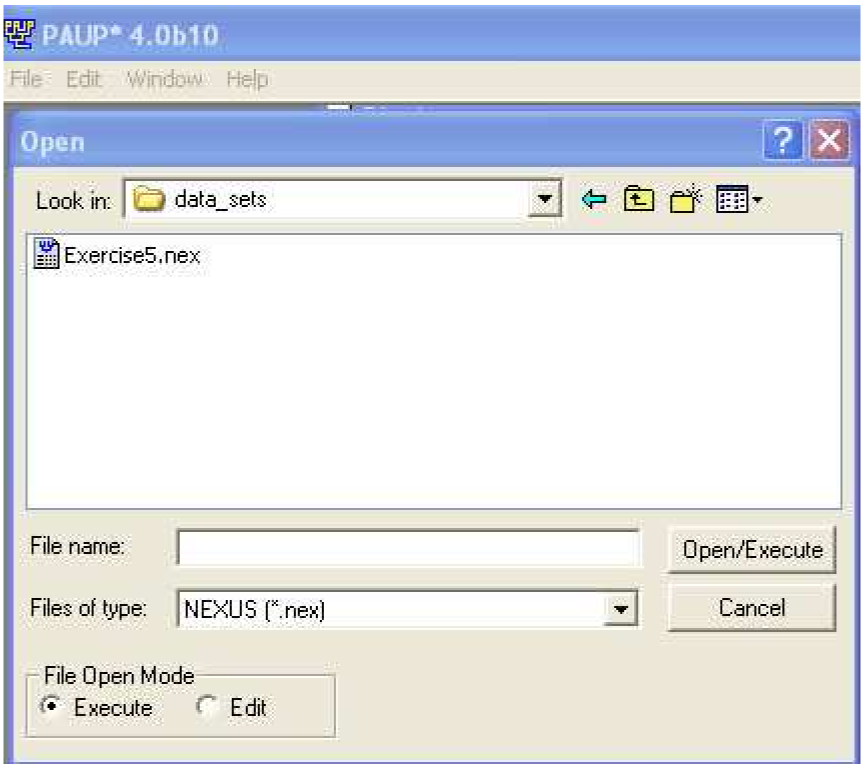
Exercise 6: Compute trees with the parsimony criterion

(BIO332 Phylogeny 2007. E.Willassen) PDF version

In this exercise we assume that you have a nexus data file prepared and will start to compute trees based on the principle of parsimony. You learn how to compute bootstrap support on clades and also how to save trees so that you can make graphics in the program TREEVIEW.

Load the data from Exercise5 into PAUP*.

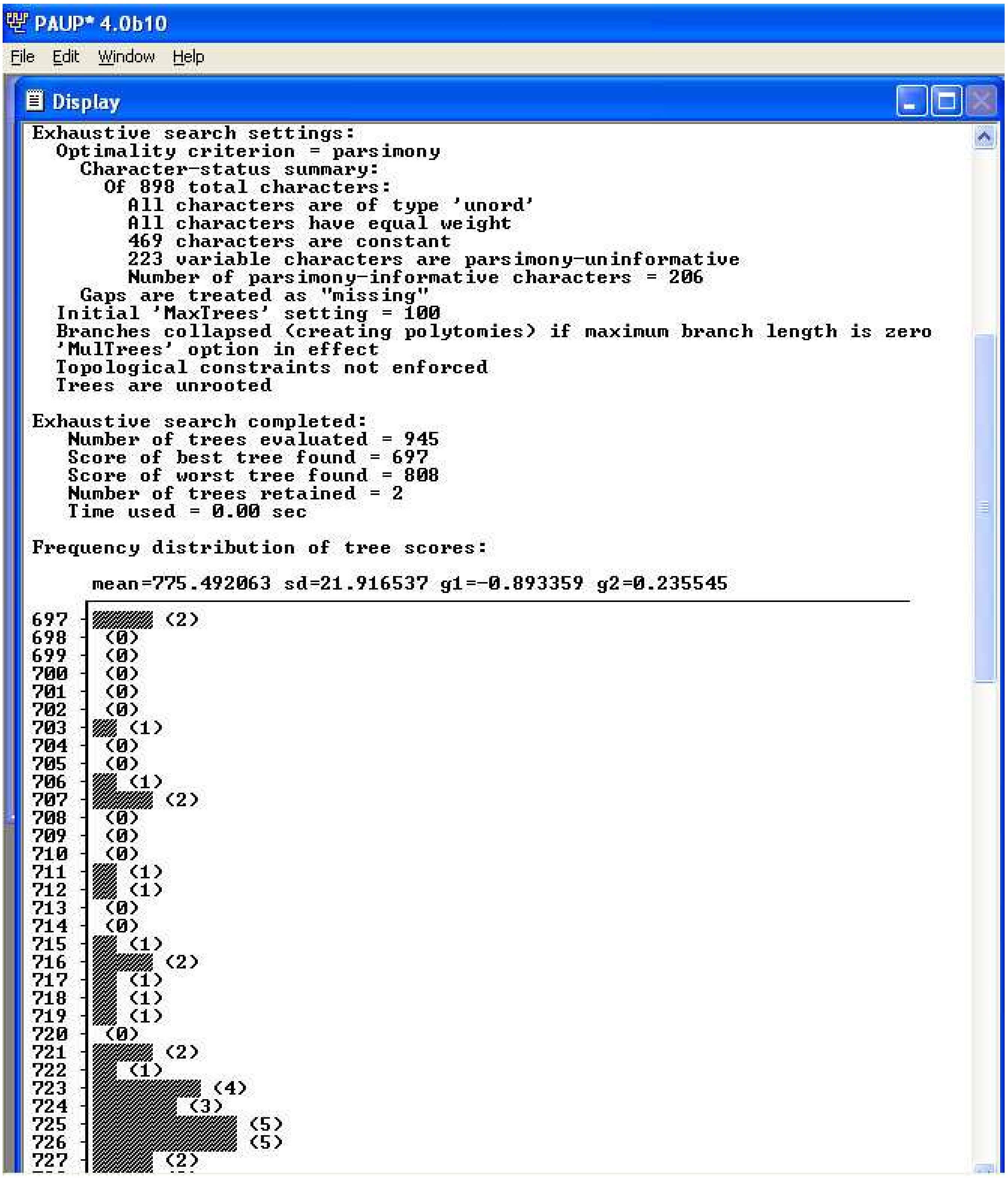


Parsimony analysis is the default in PAUP*, but if you already did calculations with other methods, you switch to parsimony with the command: set criterion = parsimony. Let us begin by exploring a command that reminds us about the number of trees that we would ideally need to examine if we wanted to find the trees with the least amount of evolutionary steps. First,delete some of the taxa in the matrix by the following command to PAUP*: del 8 - ./ only. This will temporarily exclude the range of taxa from number eight to the last one and PAUP* responds as follows in the display buffer:

```
Taxon-deletion status changed:
  5 taxa deleted
  Total number of taxa now deleted = 5
  Number of nondeleted taxa = 7
```

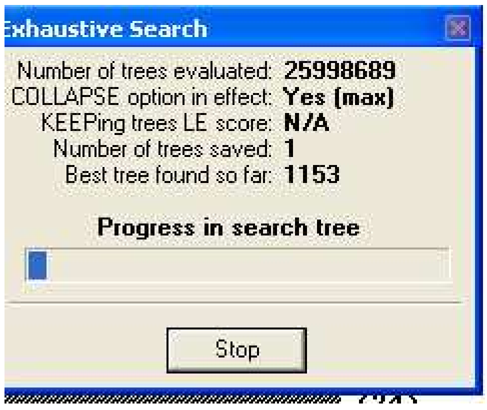
Exhaustive search

Now, write the command alltrees and execute. This results in an *exhaustive* search with length calculations over all the 945 possible trees from the seven taxa. The resulting graphical distribution of tree lengths shows that there are two most parsimonious trees with a score of 697 steps. Notice from the screen printout that the characters are unordered and that 206 of the 898 characters are parsimony informative while additionally 223 are variable, but uninformative.

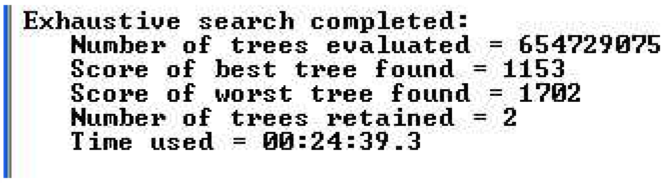


```
Exhaustive search settings:
  Optimality criterion = parsimony
    Character-status summary:
      Of 898 total characters:
        All characters are of type 'unord'
        All characters have equal weight
        469 characters are constant
        223 variable characters are parsimony-uninformative
        Number of parsimony-informative characters = 206
    Gaps are treated as "missing"
  Initial 'MaxTrees' setting = 100
  Branches collapsed (creating polytomies) if maximum branch length is zero
  'MulTrees' option in effect
  Topological constraints not enforced
  Trees are unrooted

Exhaustive search completed:
   Number of trees evaluated = 945
   Score of best tree found = 697
   Score of worst tree found = 808
   Number of trees retained = 2
   Time used = 0.00 sec

Frequency distribution of tree scores:

     mean=775.492063 sd=21.916537 g1=-0.893359 g2=0.235545
          ----------------------------------------------------------------------
697 -|▨▨▨▨▨ (2)
698 -|  (0)
699 -|  (0)
700 -|  (0)
701 -|  (0)
702 -|  (0)
703 -|▨▨ (1)
704 -|  (0)
705 -|  (0)
706 -|▨▨ (1)
707 -|▨▨▨▨▨ (2)
708 -|  (0)
709 -|  (0)
710 -|  (0)
711 -|▨▨ (1)
712 -|▨▨ (1)
713 -|  (0)
714 -|  (0)
715 -|▨▨ (1)
716 -|▨▨▨▨▨ (2)
717 -|▨▨ (1)
718 -|▨▨ (1)
719 -|▨▨ (1)
720 -|  (0)
721 -|▨▨▨▨▨ (2)
722 -|▨▨ (1)
723 -|▨▨▨▨▨▨▨▨▨ (4)
724 -|▨▨▨▨▨▨▨ (3)
725 -|▨▨▨▨▨▨▨▨▨▨▨▨ (5)
726 -|▨▨▨▨▨▨▨▨▨▨▨▨ (5)
727 -|▨▨▨▨▨ (2)
```

While the exhaustive search over 965 trees (with seven taxa) finished in less than a second, a similar search with 12 taxa (command undelete all) involves nearly 655 million unrooted trees and takes considerable more time. The status window informs about the type and status of the search.



```
Number of trees evaluated: 25998689
COLLAPSE option in effect: Yes (max)
    KEEPing trees LE score: N/A
     Number of trees saved: 1
      Best tree found so far: 1153
```

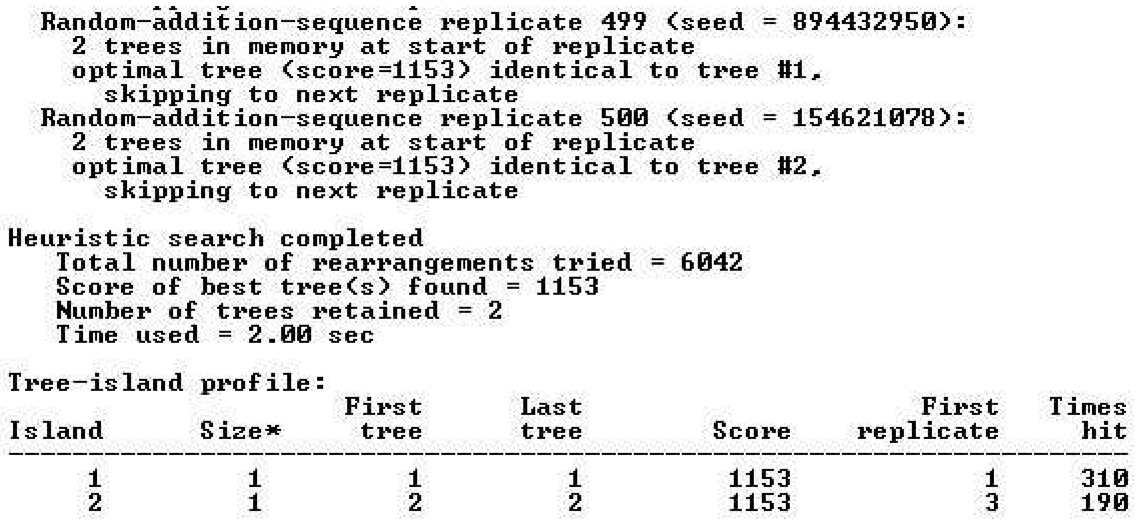


```
Exhaustive search completed:
   Number of trees evaluated = 654729075
   Score of best tree found = 1153
   Score of worst tree found = 1702
   Number of trees retained = 2
   Time used = 00:24:39.3
```

Heuristic search

With inclusion of more taxa, the exhaustive search algorithm becomes too computationally expensive at some point. We then need to seach for the shortest tree by sampling the tree space, hoping to come across the best tree in one or more of the samples. *Heuristic search* (command hs) or *branch-and-bound* (bandb) are alternative tree search options in PAUP*. Heuristic "hill climbing" algorithms may get stuck on local maxima in the landscape. In the example above, an inefficient heuristic search could come to report the tree with score 703 as the best tree, whereas according to the exhaustive search, there are two trees that are shorter: 697 steps. With a large dataset, we may increase the chance of findig the best tree by using various options with the hsearch command in PAUP*. For example, type the following in the command window: hs addseq=rand nreps=500 rstat=yes multrees=no steepest=no. When executing that you obtain 500 sampling events (nreps) with heuristic search. The first taxon in the matrix is the reference sequence, but the other taxa are added to the tree search (addseq) in random order in each replicate. That will better explore the tree space than if we added the taxa in the same order each time (addseq=asis).

We can get a report from each search by setting rstat=yes. It can be useful in some cases. We see the report format for the last two replicates in the screen print below. If PAUP* finds several equally parsimonious trees in a search replicate, it normally keeps them all because the multrees option is set to yes by default. For this example, we set multrees to no and thus obtain a report indicating that tree number 1 was found in 310 of the 500 search replicates while tree 2 was found 190 times.



```
 Random-addition-sequence replicate 499 (seed = 894432950):
   2 trees in memory at start of replicate
   optimal tree (score=1153) identical to tree #1,
     skipping to next replicate
 Random-addition-sequence replicate 500 (seed = 154621078):
   2 trees in memory at start of replicate
   optimal tree (score=1153) identical to tree #2,
     skipping to next replicate

Heuristic search completed
   Total number of rearrangements tried = 6042
   Score of best tree(s) found = 1153
   Number of trees retained = 2
   Time used = 2.00 sec

Tree-island profile:
                      First      Last                        First   Times
Island      Size*      tree      tree        Score       replicate     hit
--------------------------------------------------------------------------
     1          1         1         1         1153               1     310
     2          1         2         2         1153               3     190
```

.

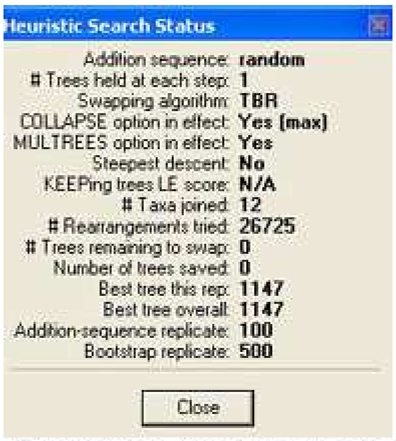
Save the trees

Save your trees with the following command: savetrees file=ex5_p_tre.tre. Then do the same again with an additional option and different filename: savetrees file=ex5_p_alttre.tre format=altnex. Open both files in Notepad / Wordpad and notice the effects of the option format=altnex.

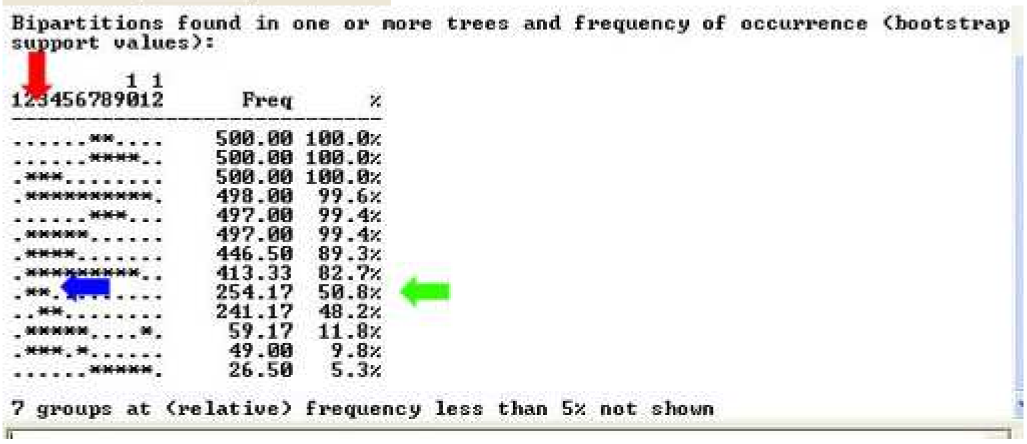
Take a closer look at the paragraph concerning *Trees* in the section "introducing PAUP*" and also study the PAUP* manual to explore the possibilities of the following commands with their options: showtrees, describetrees, savetrees.

Do a nonparametric bootstrap and save the tree with support values on branches

Depending on your data, the number of bootstrap replicates should be at least 250. The thing to keep in mind when doing a bootstrap is that you may not want PAUP* to keep hundreds of trees to swap on in each addition sequence. If you have a poorly resolved matrix, you may want to use the command set maxtrees or set multtrees=off before the bootstrap. Here is a command that does 500 bootstrap replicates and for each of those it does (default) heuristic search with 100 random additions of the taxa: boot nreps=500 / addseq=rand nreps=100; The search status window indicates that we achieved what we intended to do.

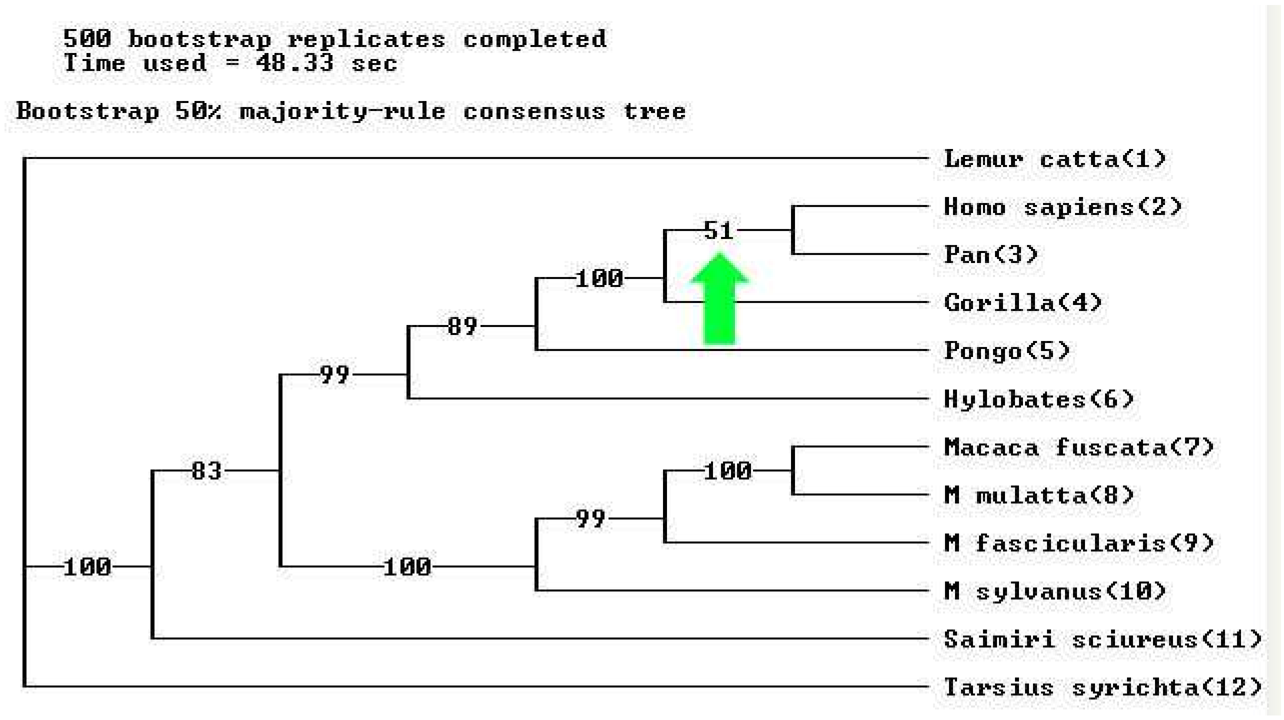


```
Heuristic Search Status
          Addition sequence: random
     # Trees held at each step: 1
         Swapping algorithm: TBR
     COLLAPSE option in effect: Yes (max)
     MULTREES option in effect: Yes
           Steepest descent: No
        KEEPing trees LE score: N/A
              # Taxa joined: 12
      # Rearrangements tried: 26725
    # Trees remaining to swap: 0
      Number of trees saved: 0
         Best tree this rep: 1147
          Best tree overall: 1147
 Addition-sequence replicate: 100
        Bootstrap replicate: 500

                Close
```



```
Bipartitions found in one or more trees and frequency of occurrence (bootstrap
support values):

           1 1
123456789012        Freq       %
---------------------------------
......**....      500.00  100.0%
......****..      500.00  100.0%
.***........      500.00  100.0%
.**********.      498.00   99.6%
......***...      497.00   99.4%
.*****......      497.00   99.4%
.****.......      446.50   89.3%
.*********..      413.33   82.7%
.**.........      254.17   50.8%
..**........      241.17   48.2%
.*****....*.       59.17   11.8%
.***.*......       49.00    9.8%
......*****.       26.50    5.3%

7 groups at (relative) frequency less than 5% not shown
```

The printout gives an opportunity to make sure we understand *bipartitions.* The blue arrow in the figure points to a partition with two taxa marked as arterisks.The red arrow above indicates that they are taxa 2 and 3 in the matrix. The green arow point to the persentage boot strap support for taxa 2 and 3 being a monphyletic group. Compare the bootstrap values in the table with those marked out on the tree.

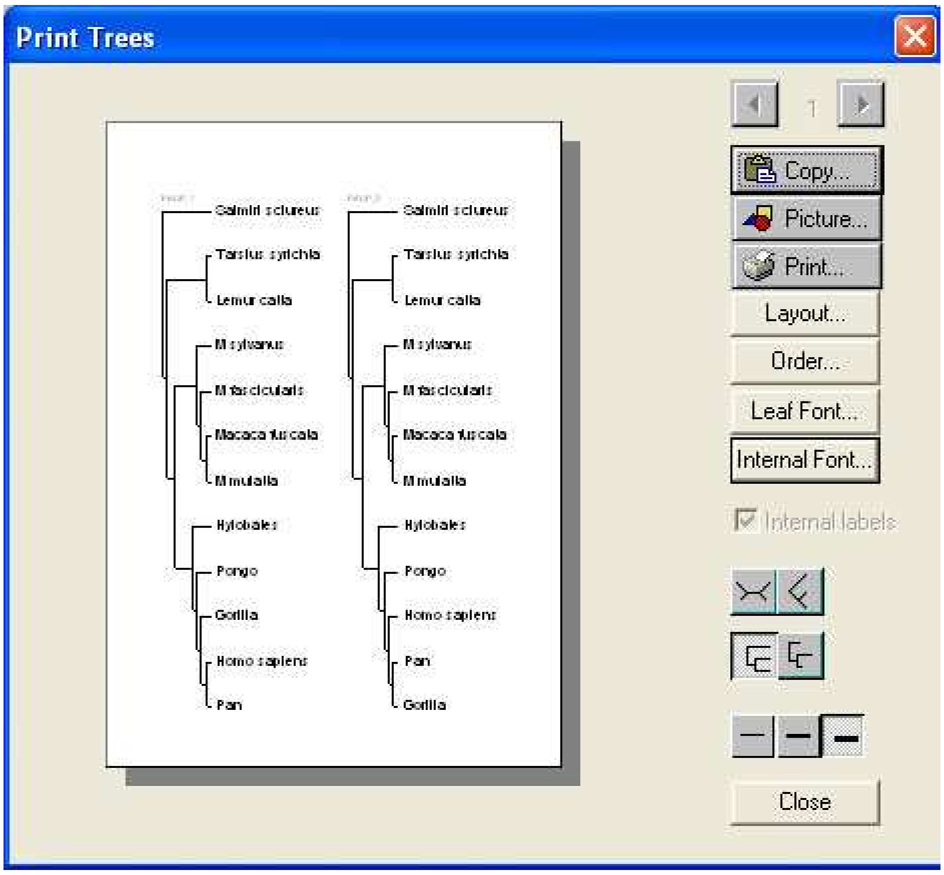


```
    500 bootstrap replicates completed
    Time used = 48.33 sec

 Bootstrap 50% majority-rule consensus tree
```

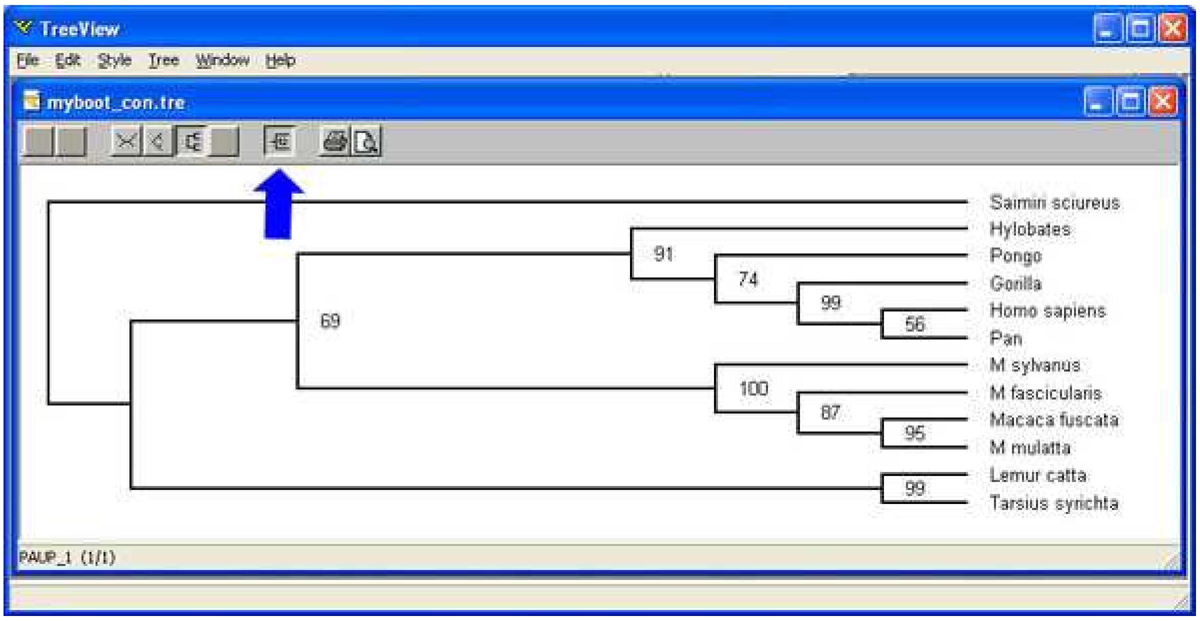
In order to save your bootstrap tree with values, type the following and execute: savetree from=1 to=1 file= myboot_con.tre savebootp=nodelabels maxdec=0 ; This saves bootstrap proportions without decimals as node labels in the tree file named myboot_con.tre.

Making nice tree graphics

Start the program **Treeview** and open your tree file ex5_p_tre.tre from the parsimony run via the **File** menu. If the tree does not contain branch lengths, you can select three different display formats by clicking the icons in the tree window. Treeview gives you the opportunity to select an outgroup and to root the tree with the selected outgroup. Notice that, by selecting the option Trees>Print Trees>Layout you can print several trees on the same page / picture.



If the tree file has Internal labels, they may also be optionally displayed if you click the icon marked with the blue arrow head. Try that by first loading the file myboot_con.tre, which has bootstrap support labels. Neat, eh!?


TreeView
File Edit Style Tree Window Help
myboot_con.tre
Saimiri sciureus
Hylobates
Pongo
Gorilla
Homo sapiens
Pan
M sylvanus
M fascicularis
Macaca fuscata
M mulatta
Lemur catta
Tarsius syrichta
91
74
99
56
69
100
87
95
99
PAUP_1 (1/1)